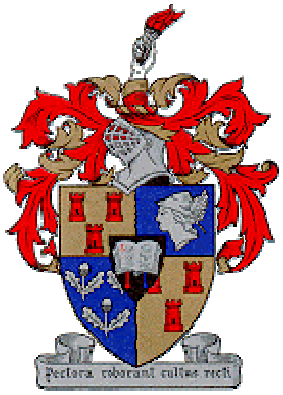
"Influence and originality in Michael Quinn Patton's *Utilization-Focused Evaluation*"

by
Motheo Precious Moleko

Thesis presented in partial fulfilment of the requirements for the degree "Master of Philosophy in Social Science Methods" at the University of Stellenbosch



Supervisor: Prof. Johann Mouton
Department of Sociology and Social Anthropology
Faculty of Arts and Social Sciences

March 2011

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining my qualification.

Signature…….. Date…….……”

Copyright © 2010 Stellenbosch University
All rights reserved

ABSTRACT

The concern of this study is primarily with insights, either acquired or original, which may be gleaned from the works of Michael Quinn Patton, especially his seminal book, "Utilization-Focused Evaluation" with regards to utilisation of evaluation. The importance of utilisation of evaluation processes and findings cannot be overemphasised. Use is critical to the success of multiple development programmes in South Africa which collectively will ultimately determine the country's accomplishment of its developmental goals. The study is an analysis of the relevant literature on utilisation of evaluation; comparing and contrasting Patton's writings with those of other scholars on this topic of utilisation. The process of understanding these insights involves a brief biography of Patton to get a glimpse into the factors which influenced the development of his ideas and to learn from them as well as from his innovative ideas. The major theme of utilization-focused approach to evaluation is utility and actual use of evaluations. The study therefore devotes special attention to these concepts in order to find out how he defines them, his motivation to focus on them and his perceptions of the purposes of evaluation. These determine his conception of the evaluation field; and his success in promoting evaluation utilisation is measured against them. The study also contains a detailed discussion of the so-called "paradigms debate" between the quantitative and qualitative paradigms in the educational and social sciences in the 1960s and 1970s. This debate was necessary to quell the division triggered by the domination of the quantitative paradigm over qualitative one in these fields and to promote understanding of the need to prioritise research questions by designing studies to answer to them instead of research problems being fitted into "one-size-fits-all" readymade designs, the study argues. The role played by Patton in turning evaluation into a profession is also discussed at length. His works on this subject include a contribution towards enhancing quality and credibility of evaluation, major input on the Standards of Excellence for Evaluation as well as assisting in the development of methods for evaluation. The inevitability of politics in evaluation is highlighted together with the importance of their ethical conduct. The debates in the American Evaluation Association are very illuminating and Patton's involvement in them is detailed in the study. They cover a range of issues which are useful to deepening of understanding of evaluation conduct and particular factors which affect use. The study ends with a critical analysis of Patton's contribution to utilisation of evaluations by programme decision makers and its significance to South Africa.

OPSOMMING

Die doel van dié studie het hoofsaaklik te doen met insigte, hetsy oorspronklik of ontleen aan ander bronne, wat afgelei kan word van die werk van Michael Quinn Patton, veral van sy invloedryke boek, "Utilization-Focused Evaluation," ten opsigte van die toepassing van program-evaluasie. Die belangrikheid van die toepassing van evalueringsprosesse en -bevindings, kan nie oorbeklemtoon word nie. Toepassing is krities tot die sukses van veelvuldige ontwikkelingsprogramme in Suid-Afrika wat gesamentlik uiteindelik sal bepaal of die land sy ontwikkelingsdoelwitte sal bereik. Die studie is 'n analise van die tersaaklike literatuur oor die toepassing van evaluering; wat Patton se werk met dié van ander skrywers oor die onderwerp van toepassing vergelyk. Die proses om dié insigte te verstaan, behels 'n kort biografie van Patton ten einde 'n vlugtige blik te kry oor die faktore wat tot die ontwikkeling van sy idees aanleiding gegee het en om van hulle, sowel as van sy innoverende idees, te leer.

Die hooftema van die toespitsing op die toepassingsbenadering vir evaluering, is sy veelsydigheid en die werklike gebruik van evaluasies. Derhalwe lê die studie hom dus veral op dié konsepte toe ten einde vas te stel hoe hy hulle gebruik het, wat sy motivering was om daarop te fokus en sy siening rondom die doelwitte van evaluasie. Dít bepaal sy begrip van die gebied van evaluasie, en sy sukses ten einde die toepassing van evaluasie te bevorder, word hierteenoor gemeet. 'n Uitvoerige bespreking oor die sogenaamde " paradigma-debat'' tussen die kwantatiewe en kwalitatiewe paradigmas in die opvoedkundige en sosiale wetenskappe van die 1960's en 1970's word ook hierin saamgevat. Verder voer die studie aan dat die debat noodsaaklik was ten einde die verdeeldheid te oorkom wat deur die oorheersing van die kwantitatiewe oor die kwalitatiewe paradigmas op dié terreine veroorsaak is, en om begrip te bevorder betreffende die behoefte om navorsingsvrae voorrang te gee deur studies te ontwerp wat dié vrae beantwoord, in plaas daarvan om navorsingsprobleme in 'n klaargemaakte een-grootte-pas-almal ontwerp te plaas.

Patton se rol om program-evaluasie in'n professie te omskep, word breedvoerig bespreek. Sy werk oor die onderwerp sluit 'n bydrae in om die gehalte en geloofwaardigheid van evaluasie te verhef, belangrike insette ten opsigte van die Standaarde van Uitnemenheid vir Evaluasie, sowel as om met die ontwikkeling van evaluasiemetodes te help.

Die onvermydelikheid van die politiek van evaluasie, tesame met die belangrikheid van etiese gedrag, kom sterk na vore. Die debatte in die Evaluasievereniging van Amerika is heel insiggewend, en Patton se betrokkenheid hierby word uitvoerig in die studie weergegee. Dit dek ’n hele reeks aangeleenthede wat nuttig is om insig ten opsigte van die bestuur van evaluasie te versterk, veral faktore wat die gebruikswyse beïnvloed. Die studie sluit af met ’n kritiese analise van Patton se bydrae tot die toepassing van evaluasies, soos dit deur programbesluitnemers gebruik word, en sy betekenis vir Suid-Afrika word ook ondersoek.

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Prof Mouton for the completion of this study. I am very grateful for the invaluable guidance and support he gave me throughout this study.

I thank Marthie van Niekerk for patiently helping me every time the need arose.

I am particularly thankful to my parents, Mamotheo and Tefelo Moleko for their encouragements.

To my daughters, Tshiya and Lehakwe Moleko, thank you for the motivation to keep on studying when I felt like giving up. To my niece, Aseza for the support she gave me.

Special gratitude goes to my partner, Mponeng Moleko for the unconditional support, encouragement and sacrifices she made from the beginning until the completion of this study.

TABLE OF CONTENTS

ABSTRACT ……..iii
OPSOMMING……...iv
ACKNOWLEDGEMENTS……...vi

1. CHAPTER ONE: INTRODUCTION……1

1.1 Introduction……...1
1.2 Motivation for the study……....1
1.3 The aims of the study……2
1.4 Methodology……4
1.5 Limitations…….....5
1.6 The outline of the study……5
2. CHAPTER TWO: THE DEVELOPMENT OF PATTON'S IDEAS...11

2.1 Introduction……11
2.2 Biographical sketch……11
2.3 Sources of influence on Patton's ideas……12
2.3.1 Training in positivist quantitative methodology……...12
2.3.2 Interest in knowledge diffusion and utilisation……14
2.3.3 Exposure to qualitative research and Egon Guba's reaffirmation……18
2.3.4 Influence of John Dewey's progressive education and pragmatism……23
2.4 Conclusion……...25

3. CHAPTER THREE: PATTON'S NOTION OF UTILISATION…….27

3.1 Introduction……..27
3.2 The meaning (s) of "evaluation research"……27
3.3 Reasons for Patton's focus on use……32
3.3.1 Unmet expectations created by positivism, modern science and Rationality……32
3.3.2 The utilisation crisis……..33
3.4 Barriers to use……..34
3.5 The importance of use……..37
3.5.1 Typology of utilisation……..37
3.5.2 Purposes of evaluation……...44

3.5.2.1 Formative evaluation……44
3.5.2.2 Summative evaluation……..45
3.5.2.3 Knowledge generation……..46
3.5.2.4 Developmental purpose……....47
3.5.3 Beneficiaries and spheres of evaluation utilisation…….49
3.6 Misutilisation…….50
3.7 Conclusion…….56

4. CHAPTER FOUR: PATTON'S PARTICIPATION IN THE PARADIGMS DEBATE…….57

4.1 Introduction……...57
4.2 Early reaction to the dominance of quantitative paradigm…….....57
4.3 Opposition of objectivity to subjectivity……...62
4.4 Debates about designs……...64
4.5 Focus of evaluation……...66
4.6 Contextual inimitability and generalisation……..67
4.7 Evaluation for Whom and for What? 69
4.8 Paradigm of choices……..71
4.9 Quality and credibility in qualitative evaluation studies..……72
4.10 Conclusion……..75

5. CHAPTER FIVE: PROFESSIONALISATION AND POLITICS OF EVALUATION PRACTICE……76

5.1 Introduction……..76
5.2 Patton's vision of a professional evaluation discipline……77
5.2.1 Commitment to utility……77
5.2.2 Guidance on enhancement of quality and credibility in qualitative analysis……..78
5.3 Credibility of the evaluator……...82
5.3.1 Evaluator training, experience and preparation……..82
5.3.2 Evaluator and evaluation effects…….....83
5.4 Philosophical beliefs……85
5.5 Case studies as a tool for professional development of evaluation……..86

5.6 Barriers and challenges to the professional development of evaluation…….89
5.7 Comment on Standards of Excellence for Evaluation Practice……...91
5.8 The relationship between evaluation and other disciplines…….92
5.9 The Politics of evaluation……..102
5.9.1 The political inherency of evaluation…….102
5.9.2 Evaluation and democracy……..104
5.9.2.1 Factors that affect democracy in evaluation..……....108
5.9.2.2 Effects of democratic participation in evaluation……..110
5.10 Management of politics in evaluation……..113
5.11 Ethical issues in qualitative research and evaluation……...115
5.12 Conclusion……....118

6. CHAPTER SIX: PATTON'S INVOLVEMENT IN THE DEBATES WITHIN THE AMERICAN EVALUATION ASSOCIATION……119

6.1 Introduction……....119
6.2 The Patton-Weiss debate on the limits to evaluation use...……....119
6.3 Patton's debate with Scriven on the formative-summative distinction……..143
6.4 Patton challenges Fetterman on empowerment evaluation……....155
6.5 Kirkhart's dispute with Patton's focus on 'utilisation'……...161
6.6 Henry's discussion of Patton and the notion of "social betterment"……..168

6.7 Critiques of other evaluation approaches……...172
6.7.1 An assessment of the theory-driven approach to validity……...172
6.7.2 An appraisal of feminist evaluation……....174
6.7.3 The review of Egon Guba and Yvonna Lincoln naturalistic dogmatism....181
6.8 Conclusion……..182

7. CHAPTER SEVEN: CONCLUSION……..184

Bibliography……190

List of figures

Figure 3.1……..43
Figure 3.2……..54
Figure 3.3……..55

CHAPTER ONE: INTRODUCTION

1.1 Introduction

This chapter comprises the rationale of this study. It outlines the reasons I chose the topic of this study as well as what I hope to accomplish through it. The methodology also forms part of this chapter and it describes the processes followed to accomplish the aims of the study. The last section is the outline of the study which briefly delineates the contents of the remainder of the chapters.

1.2 Motivation for the study

Governments and non-governmental organisations around the globe have identified programmes as one of the major tools in driving their quest to alleviate and where possible eradicate human suffering. Given the centrality of programmes in accomplishing these noble gaols and the amount of resources invested in them, it is imperative that their activities be evaluated. Programme evaluation emerged out of a need to address a range of issues concerning programme effectiveness, programme improvement, informing decision making in programmes or organisations in which programmes are located and a host of other purposes. Evaluation has been used for wide range of purposes. Depending on the approach, evaluation has been used to improve the efficiency of the programme, to generate knowledge of programme effectiveness, to determine whether or not the programme is operating as it should and so on. Different approaches to evaluation have had varying degrees of success in their intended goals depending on the rigour with which they were applied and their attention to utility but the utilisation crisis in the 1960s and the 1970s in the United States of America was symptomatic of a malady in the manner in which evaluation was practised. It functioned mainly as an oversight mechanism with little attention to utility in its heavy emphasis on judgement of merit or worth. It was against that backdrop that Michael Quinn Patton wrote "Utilization-Focused Evaluation" with the mission to turn around the situation in the evaluation field of practice. He writes about the book that "the approach presented here is derived from the observation that much of what passes as evaluation is not very useful; *that evaluation ought to be useful*; and therefore that something different must be done if evaluation *is* to be useful. This book suggests what it is that should be done differently and how it should be done" (Patton, 1986: 7). It is undeniable that evaluations cannot produce desired outcomes if they are not utilised and considering the import of evaluation information in improving human conditions, it

is very critical that it be used. Indeed, the evaluation landscape has changed since the first publication of the book in 1978. The prominence of this approach and its emphasis on 'intended use by intended users' has principally shifted attention away from the foci of other approaches before it to utilisation. The choice of Patton was also motivated by the fact that he was involved in the inaugural conference of the African Evaluation Association (AFREA) held in Nairobi in 1999 where he delivered a series of lectures in evaluation. The fact that those lectures were compiled into a guide to evaluation in Africa (Chaiban and Patel, 1999: 1) shows endorsement by AFREA.

1.3 The aims of the study

The name of the approach "Utilization-focused evaluation" is inviting in its emphasis on use. Reason dictates that anything has to be used in order for it to be useful, that is, the value of evaluation depends on its use. The main purpose of this thesis is to determine the degree of originality of ideas contained in utilization-focused evaluation theory and the extent to which they have contributed towards extricating evaluation from the so-called "utilisation crisis" in the USA in the 1960s and 1970s by contributing towards promotion of a culture of evaluation utilisation as well as making evaluations useful. I also intend to make a contribution in the country - South Africa; by playing a role in a task to build a repository of information on evaluation utilisation. To the best of my knowledge there is no review on the use of evaluations of programmes meant to effect social, economic and cultural changes in South Africa. After 1994, the South African government embarked on a broad policy programme known as the Reconstruction and Development Programme (RDP). This programme was an umbrella for multiple projects aimed at ameliorating poverty and addressing various basic shortfalls in social services across the country. The multi-strand programme is composed of a multitude of programmes whose purpose is to foster growth of the country's economy and to link it to development through reconstruction and redistribution of the proceeds of economic growth. The present government has also pledged its commitment to reviving RDP as a vehicle to social upliftment. However, many of the programmes which have been conceived over the last fifteen years have been riddled with problems associated with incompetence in implementation on the part of programme staff or graft by government officials, problems which might have been picked up by evaluators. For example, there are numerous unfinished housing projects throughout the country and the government finds itself in an untenable position whereby it has to dip deep into the national treasury's coffers to finance their

rebuilding. It is not far-fetched to suggest that some of the problems the government faces today are related to the incapacity of its employees to implement its programmes/projects as well as their monitoring and evaluation. In looking at elements of evaluation utilisation, this study is a small but necessary contribution to the literature on evaluation utilisation in South Africa which may be useful to the overall development in the country.

Evaluation functions as a form of an instrument of compliance with purposes of programmes or policies in its judgemental nature and its claims to empiricism, despite its lack of enforcement mechanism. However, the necessity for such an emphasis on utilisation also implies that something is amiss in evaluation. The rational assumption is that evaluation is conducted to provide empirical information that is used to inform decision making processes for improvement, judgement of merit or worth, knowledge generation and other purposes of evaluation. But Patton declares that much of evaluation was not useful at the time of the writing of this book (1978) and that the book suggests new ways of carrying out evaluations in order make them useful. This is the statement that captured my attention, especially in the light of multitudes of programmes/projects which have failed in South Africa. One of the aims of this thesis is to highlight practices which are instrumental in making evaluations useful. I explore the extent to which the ideas about utilisation in the book overlap and break with the previous practice in evaluation and the reasons they are included in the book – how they resolve the problem of non-utilisation of evaluation recommendations and what can be learned from them. Influence presumes a social construction and constructivist view that there were elements/sources of information and experiences either from within or outside evaluation theory and practice at the time when Patton wrote the book which informed his approach to evaluation. Special attention is also devoted to exploring new ideas incorporated into existing theory on utilisation and their impact on use of evaluation processes and findings. An overall summation of the strengths and limitations of utilisation-focused evaluation are also delineated as well as its appropriateness for developing countries such as South Africa.

The other aim is to learn from U-FE and highlight those areas about it which may assist SA in avoiding the situation that the United States of America found itself in during the utilisation crisis. This is critical in the light of the country's drive for redress of social

and economic differentiation through programmes. This is related to the role U-FE can play in building capacity and promoting learning in programmes and organisations.

1.4 Methodology

It is essential that any study be conducted employing a methodology responsive to the research question; in this study the relevant and appropriate methodology for its conduct is literature study/review, which I have used. The review is primarily an assessment of Michael Quinn Patton's seminal work "Utilization-Focused Evaluation" beginning with its first publication in 1978 to the fourth edition of 2008 in order to establish the extent innovation in his ideas. During the period spanning the years from 1975 to 2008, Patton wrote numerous articles and books, ideas of which invariably found their way into following editions of the book in question – pertinent works to utilisation among those have been critically assessed and included. The review summarises Patton's ideas and organises them into themes with a view to assessing his motivation in writing this book, how his ideas have developed over time and the degree to which his thoughts are different from the existing stock ideas. This enables us to discern whether there is any fundamental departure from common practice in evaluation in what he offers in his approach to resolve the issue of evaluation under-utilisation or non-utilisation.

Part of the process of identifying his contribution to utilisation of evaluation has been the investigative reading the critiques and commentaries of his works. In order to determine the influence and originality of Patton's ideas, particular attention is paid to traceable influences that might have led to its conception, and its groundbreaking aspects, its strengths and weaknesses with regard to various aspects of interest to programme evaluation. For instance, its input to the body of scientific knowledge, theory development and most importantly, its contribution towards the promotion of evaluation utilisation. This has been helpful in demonstrating the impact of the utilization-focused approach on the profession of evaluation. The reasons put forward in these critiques both in support of the approach and in opposition to it are very illuminating and useful to the overall purpose of this study.

1.5 Limitations

The limitation of this study is that I have been unable to access literature in utilization-focused evaluation with reference to Africa, especially South Africa. I intended to use this literature as case studies whereby selected works by certain evaluation practitioners who have adopted utilization-focused evaluation and applied it in their studies in various fields and states in Africa are assessed. The idea behind this is to find out the applicability of Patton's approach, that is, how its theory translates into practice, its utility with regards to promoting utilisation and its limitations. The significance of this exercise is its efficacy in providing information for weighing advantages and disadvantages of utilization-focused evaluation. This information is also valuable in establishing the appropriateness of this approach to the South African context.

1.6 The outline of the study

The introduction constitutes the first chapter. In it, I discuss the rational of the study which is composed of my motivation to focus attention on Michael Quinn Patton's "Utilization-Focused Evaluation". The aims of the study and the methodology employed in it also comprise part of this chapter. Thereafter, I draw the outline of this study which follows below.

In the second chapter I map out the development of Patton's ideas that culminated in his seminal book "Utilization-Focused Evaluation" in 1978. I begin with a brief biography in order to illustrate his professional standing and the high regard with which he is held in the evaluation research and practice community. This is evident in the numerous awards bestowed on him over a period spanning four decades as a professional evaluator, consultant and writer. I then trace the development of his ideas from a coincidental chain of events that saw him register as a graduate student in sociology at the University of Wisconsin. It was during that training as a sociology student at graduate and postgraduate levels that his ideas began to take shape – sociology anchored him in quantitative methods, theory construction, sociology of knowledge and other aspects that had an impact on his thinking. Moreover, sociology introduced him to the works of Everett. M. Rogers on diffusion of knowledge. That proved to be critical in so far as it registered in Patton the importance of the role of stakeholders in evaluation research. That is what informs utilization-focused evaluation with its emphasis on intended users.

I have included Patton's sojourn at the University of North Dakota's New School for Behavioral Studies in Education which marked the turning point in his methodological outlook to encompass qualitative evaluation. The monograph, "Alternative Evaluation Research Paradigm" (1975), was a result of the influence; it consists of a fervent plea for qualitative approach. It is during his time at North Dakota that he met Egon Guba who had a profound effect on him by reaffirming his ideas on qualitative evaluation. By that time (1977), Patton was firmly involved in the so-called "paradigms debate", throwing his weight behind the qualitative approach. At the same time, the influence of the pragmatist education scholar, John Dewey had impressed upon him the necessity of tolerance of diverse ideas. That is evident in utilization-focused evaluation approach's accommodation of multi-method designs which are adapted to research questions and situational factors instead of designs determining research questions.

Chapter three is a discussion of Patton's application of the notions of *evaluation research*, *utility* and *use* in evaluation research which form the basis of utilization-focused evaluation. I found it necessary to give a definition of evaluation research from Patton's perspective, a view that evolved over time. A definition becomes imperative in the light of multiple and sometimes contrasting definitions by evaluation practitioners and scholars. The understanding of Patton's perception of evaluation research will enable readers to understand his writings. I therefore map the evolution of his understanding of evaluation research from the 1970s and thereafter. Similarly, I found it necessary to unpack how he employs *utility* and *use* because of their centrality to his approach. Patton employs *utility* in different ways and also appears to have a different understanding of *use* from other evaluation practitioners who were concerned about the fact that results of many evaluations were not used. According to him, the definition of use might have contributed to perceptions of little utilisation of evaluation.

A discussion of positivism is also in chapter three because of its influence at the time. A major proposition in positivism is that scientific methods (those used in natural sciences) should be applied to all areas of investigation, including social sciences. Positivism holds the view that the logic of inquiry is supposed to be the same across all sciences, a factor that led to the paradigms debate in social sciences in the 1960s and 1970s. At the time of his writing of "Utilisation–Focused Evaluation" (1978), Patton was aware of the evident influence of positivism on the die-hard proponents of the scientific method (quantitative paradigm) and its stranglehold on methods of social

investigation. The positivist's optimism about the promise of modern science and the creation of a better world and rationality presupposed that decision makers would logically apply findings of properly conducted evaluations. Patton's approach to evaluation was/is not only a reaction to the scientific method (equated with "quantitative"), but also a rejection of positivism which actually underpinned the scientific method.

I have included a discussion of the so-called "utilisation crisis" in this same chapter. This "crisis" referred to the "absurdity" of evaluators carrying out evaluations that were little used or not used at all as well as clients pumping in financial resources into a practice that did not justify their investments. Several authors (e.g., Weiss, Chelimsky, Freeman and Rossi *etc*.) have written about the utilisation crisis but it is included here because it led directly to Patton's focus on use. His choice for utilisation-focused evaluation was a result of a reaction by Patton to a question he was asked by a client about prerequisites to improve evaluation use. According to Patton, evaluation use or lack thereof was linked to the employment of non-empirical approaches to evaluation such as charity inclination and pork-barrel approaches as well as evaluator pitfalls that hampered utilisation even when evaluation approaches were data-based. What follows is the importance of use as manifested in the multiple types of use and an extended discussion of process use because of interest in and issues around it, purposes of use, beneficiaries and spheres of evaluation utilisation as propounded by Patton (beneficiaries of evaluation and spheres help to elucidate the importance of use by indicating the scope and levels of users - if it can be used by such a range of people and groupings, then it is important) and misuse.

In the fourth chapter I provide the backdrop against which most of evaluation research was conducted during the 1960s and 1970s and which ultimately sparked the paradigms debate between quantitative and qualitative approaches to evaluation studies. The employment of experimental and quasi-experimental designs in social sciences and its applied versions such as "classic" evaluation theory (Campbell and Stanley) has its roots in positivism. The rise of the experimental method coincided with the parallel ascendance of logical positivism in the middle of the twentieth century. In social sciences research experimental domination manifested itself in the works of proponents like Campbell and Stanley (1963), Cook and Campbell (1979) who were prominent proponents of experimental designs as *the* models of research, while rating methods

related to the qualitative paradigm very poorly. These theorists and a host of others firmly believed in the unassailable importance of experiments in determining cause and effect and establishing general laws for human behaviour.

Proponents of this approach to evaluation believe that experiments and randomization ensure objectivity. Exponents such as Campbell viewed social interventions as social experiments and as such could be studied employing experimental approach.

The professionalization of evaluation practice and its politics are dealt with in the fifth chapter. In the early 1980s, the paradigms debate had raged past its apex and there was a shift away from hard-line postures explicated by adherents of quantitative and qualitative paradigms. There was an increasing convergence of opinions from the opposing factions on the necessity to match research methods to research questions. The Standards of Excellence for Evaluation were published in 1981 and they regulated the manner in which evaluations were to be conducted by evaluators irrespective of their paradigmatic affiliation. The Evaluation Research Society and the Evaluation Network also merged in 1984 to form the American Evaluation Association in a step that envisaged a yearning towards unity of evaluation research and practice. However, certain questions which cast evaluation research in dim light were still outstanding. For instance, besides a considerable growth in its body of literature and a number of full-time practitioners, evaluation research still remained largely a step child in the service of other professions without an independently acknowledged existence. This chapter is therefore, concerned with Patton's contribution in the challenge of extricating evaluation research from the shadows of other professions and turning it into a fully fledged professional discipline. Patton suggests that in order for it to move forward and earn respect of other professions, it had to have a vision that promotes quality which in turn require sound training in research methods and an abundance of technical expertise. While methodological competence and technical skills are critical, utility was emphasised as the vehicle to evaluation's professional acceptance. Commitment to use received a major boost with the publication of Standards of Excellence for Evaluation (1981) which put utility at the centre of evaluation practice. There was also a realisation that evaluation needed new methods and approaches in order for it to meet its increased needs necessitated by its growth, hence propositions by evaluators like Nick Smith and Patton to look beyond their practice for new approaches and adapt them to their field to enable them to match methods to the needs of varied disciplines that evaluation covers.

Different disciplines that have similarities with evaluation are discussed in this regard. Patton encourages evaluators to commit to the profession by taking forward what he perceives as a roadmap to the success of the profession, namely; a vision that promotes utility, quality products and processes and skilled, trained evaluators.

In chapter six I discuss some of the debates Patton was involved in within the American Evaluation Association, starting with the ongoing but illuminating debate between Patton and Carol Weiss. The debate was sparked by Weiss' presentation in the second meeting of the AEA in Boston, October 1987 during Patton's presidency of the association. I deemed it necessary to paint a vivid picture of the issues involved, hence a detailed digestion of the authors' first presentations, that is, Weiss speech at the conference and Patton's response to it. Patton's response to Weiss' presentation is not necessarily a point by point refutation of the latter's reasons for evaluation's 'indifferent success', instead it is a presentation of an alternative vision which emphasises planned and direct use. The nature of the exchanges does not allow a detailed denunciation of each other's arguments in so far as they are not direct responses to the points raised, but rather, different visions whose scope converge and diverge on certain issues. In this instance, convergence and divergence refer to common and disparate topics covered. Nonetheless, some points raised by Weiss are covered in Patton's vision and I highlight them as they are presented. In a speech that was later to appear in the article under the title**,** "Evaluation for decisions: *Is anybody there? Does anybody care?" Evaluation Practice***,** 9 (1): 5-19 (1988)*,* Carol H. Weiss presented her views on evaluation utilisation. The paper was to become the first in a series of exchanges between her and Patton. In that conference, Weiss gave a presentation on the obstacles to evaluation use and the reasons traditional advice to evaluation resulted in 'indifferent success'. Patton felt that this was an attack on utilization-focused evaluation. Patton's debate with Weiss is longer and more heated than his exchanges with other scholars probably because it dwells on utilisation, a topic close to Patton's heart and the numerous reasons put forward for and against direct use.

Patton's debate with Scriven was a response to the latter's definition of formative and summative evaluations in "Beyond Formative and Summative Evaluation" (1981). Patton felt that the definition was restrictive about evaluation purposes in the light of the growth of the profession. Their exchanges reveal glimpses of their postures on the paradigms debate. The last debate involves Patton exchanges with Fetterman on

empowerment evaluation. While the exchanges with Weiss, Scriven and Fetterman were initiated by Patton's response to the two authors' works, the debates with Henry and Kirkhart were reactions to Patton's writings. Both authors do not feel comfortable about the pervasiveness of use or utilisation in evaluation literature as spearheaded by utilization-focused evaluation. Henry stresses social betterment as the ultimate purpose of evaluations while Kirkhart points out the limitations of use in covering process-based influence. I discuss the contributions of Henry and Kirkhart at length to provide the background for their arguments because they are premised on their theories of social betterment and influence respectively. Finally, I have added short critiques of theory-driven, feminist and fourth generation evaluations by Patton.

The seventh and final chapter is devoted a concluding discussion of the value and legacy of Michael Patton's work in the field of evaluation studies, limitations of utilization-focused evaluation and what can be learned from them to the South African context to improve evaluation utilisation.

CHAPTER TWO

THE DEVELOPMENT OF PATTON'S IDEAS

2.1 Introduction

This chapter is devoted to outlining the development of Patton's ideas which culminated in his writing of the book "Utilization-Focused Evaluation" in 1978. It begins with a short biography with the purpose to demonstrate influences that shaped his thinking. I then trace the development of his ideas from a coincidental succession of events that saw him register as a graduate student in sociology at the University of Wisconsin. It was during that training as a sociology student at graduate and postgraduate levels that his ideas began to take shape – sociology anchored him in quantitative methods, theory construction, sociology of knowledge and other aspects that had an impact on his thinking. Moreover, sociology introduced him to the works of Everett. M. Rogers on diffusion of knowledge. That proved to be critical in so far as it registered in Patton the importance of the role of stakeholders in evaluation research. That is what informs Utilization-Focused Evaluation with its emphasis on intended users.

2.2 Biographical sketch

Michael Quinn Patton began his graduate studies at the University of Wisconsin, studying "highly quantitative sociology" and completed his doctoral degree from the same university. He did fieldwork for his doctoral dissertation with the North Dakota Study Group on Evaluation under the tutelage of Vito Perrone conducting evaluation of open classrooms all over North Dakota where he came to grips with qualitative evaluation research and began to appreciate it. After receiving his doctorate in Sociology from the University of Wisconsin, he spent 18 years on the faculty of the University of Minnesota (1973-1991), including five years as Director of the Minnesota Center for Social Research and ten years with the Minnesota Extension Service. During this time he was elected (1987) President of the American Evaluation Association (AEA), a position he held the following year.

Patton has received several awards for outstanding performance in the field. These include the University of Minnesota *Morse Amoco Award* for exceptional teaching, the *Alva and Gunner Myrdal Award* from Evaluation Research Society for "outstanding contributions to evaluation use and practice" and *Paul F. Lazarfeld Award* for lifetime

contributions to evaluation theory from AEA. He also became a faculty member of the Union Institute Graduate School which specialized in individually designed, non-residential, non-traditional and interdisciplinary doctoral programmes. In 2001, Patton was awarded the *Lester F. Ward Award* by the Society for Applied Sociology for extraordinary contributions to Applied Sociology. He also became a member of the Editorial Board of Science Communication.

At present, he is a generalist consultant in multiple disciplines that attempt to alleviate human deprivation that include evaluation and organisational development. For instance, he provides regular training for the International Program for Development Evaluation Training under the auspices of the World Bank, The Evaluators' Institute as well as AEA pre-conference professional development courses. Patton has written five books on research evaluation, namely: *Utilization-Focused Evaluation* (1978), *Qualitative Research and Evaluation Methods* (1980), *Practical Evaluation* (1982), *Creative Evaluation* (1987) and co-authored *Getting to Maybe: How the World is Changed* (2007) and numerous scholarly articles across a range of disciplines.

2.3 Sources of influence on Patton's ideas

2.3.1 Training in positivist quantitative methodology

One of the earliest influences in Patton's world outlook, which also served as a door to his entry to evaluation, comes from sociology. His choice of sociology followed a chain of events which commenced from an unexpected quarter – the Beatles, a 1960s rock music band. He writes that: "In paying homage to root influences, I have to include the Beatles. Yes, the 60s rock group. You see, they are responsible for my becoming a sociologist, and sociology has shaped my world view about evaluation use" (Patton, 2004: 279). He recounts how he was mesmerised by 'the hysterical screaming, unrestrained adulation, frenzied dancing and occasional fainting' during the group's appearance on the Ed Sullivan Show. He thereafter wrote an English course essay about the experience, only to be advised by the unnamed course professor who felt his fascination with the experience, to venture into sociology to quench his thirst for answers to his enthralment – advice which he followed. That had far-reaching consequences in his early academic perspective on paradigms as he acknowledges:

> "Sociology gave me a solid grounding in quantitative methods, theory construction, philosophy of science, and sociology of

> knowledge, all of which have influenced my evaluation thinking and practice" (Patton, 2004: 280).

This point is further underlined when he states that he is a product of an institution, University of Wisconsin, where quantitative methods were highly esteemed during his time as a student and where his 'graduate training had been entirely quantitative' (Patton 2008: 319). This is not surprising given the fact that he grew up at a time when positivism, particularly logical positivism held sway. The driving force behind logical positivism was the formation of a scientific society comprised of all scientists from different scientific fields. However, unity in logical positivism came at a high cost for other sciences owing to the fact that their laws had to play second fiddle to those of physics. For instance, Rudolf Carnap's idea of unity is reported thus: "He did believe, however, that the various laws discovered in each of the sciences could eventually be reduced to one another, and all laws reduced ultimately to those of physics" (Babbie, E. and Mouton, J. 2001: 24). It was therefore not a coincidence that Patton studied quantitative methodology because great scholars of social science were influenced by this tradition. Among them are luminaries such as Campbell and Stanley who propagated the experimental model in pursuit of identifying causal relationships (internal validity) in the deductive-nomological style as the benchmark of evaluation studies. The stranglehold of the quantitative paradigm on social research in the 1950s and 1960s is further signified by Bernstein and Freeman's coding of evaluation quality variables in their 1974 study whereby they coded all quantitative variables higher than qualitative ones in quality (Campbell and Stanley [1974] as cited by Patton, 1975: 9-10). They then made a rather lengthy list of scholars whose work was quantitative to justify their standpoint. While that may not be sufficient justification, it is significant in portraying the background against which Patton acquired his training in methodology.

It was during his time as a student at the University of Wisconsin in the 1960s that he was introduced to organisational sociology and learnt about change management from the literature of the likes of Etzioni, Hage and others. The fundamental idea that captured Patton's attention in the theorists' discussion of conflict is the pronouncement that it is not always the substance of change itself that is the source of resistance to change, but rather how the transition from one way of doing things to another is managed. In one extreme, the tendency is for the management to exclude other stakeholders, thus causing disaffection among the sidelined. On the other end of the pendulum, there is excessive engagement which prolongs the process to an extent that

when change occurs, it is almost difficult to tell the influence of all stakeholders. Preference is given to a middle path between the two extremes. This influence is evident in Patton's approach to stakeholder involvement in utilization-focused evaluation whereby attempts are made to tackle conflict in a manner that will satisfy all stakeholders, particularly intended users of evaluation.

Patton declares that one of the sociological specialisations that had influence on him is the theory of power as propounded by theorists like Michael Crozier. It taught him about power relationships among individuals and groups by providing him with "a basis for understanding how knowledge is power and led me to the following premise: "Use of evaluation will occur in direct proportion to its power-enhancing capability" (Patton, 2004: 281). In "The Bureaucratic Phenomenon", Crozier explores the relationship between power and uncertainty. He contends that management employ evaluation as an instrument of control in the sense that its findings serve to reduce certainty for action in specific departments. By furnishing information that can be utilised for prediction, evaluation also provides the possibility for control because some rules and regulations with predictable outcomes can be put in place. The ability to predict stakeholders' behaviour is control over them and the only way they can wrest back power is through exploitation of areas of uncertainty where outcomes are not predictable. Therefore organisational relations are power struggles among individuals and groups to gain ascendancy over others.

Another issue which might have appealed to Patton is Crozier's attack on centralisation of decision making in the upper echelons of organisations. Crozier found fault with the process in which people without direct knowledge of issues on the ground wield power of influence over those with direct experience. Again, those on the ground resort to areas of uncertainty to assert their importance and power, thus creating something akin to parallel centres of power and concomitant instability. The centrality of intended users in utilization-focused evaluation is an attempt by Patton to place decision-making powers in the hands of stakeholders evaluation is intended for.

2.3.2 Interest in knowledge diffusion and utilisation

Patton (2004) maps out the development of his ideas which later became the basis of utilization-focused evaluation from his time as a Peace Corps volunteer among the Gourma people in the eastern part of Burkina Faso in the 1960s. He declares that it is

while he was there that he learned empathy, interpersonal and organising skills as well as pragmatism which became an integral part of his approach to evaluation. He writes:

> "I learned how to figure out what someone cared about, how to bring people together to identify shared interests, and how to match initiatives and resources to those shared interests. I learned to ground my change efforts in the perspectives, values, and interests of those with whom I worked, the indigenous people who were there before I came and would be there after I left. I learned to appreciate and honor local villagers and farmers as the primary stakeholders in change and to see my role as facilitating their actions, not letting my interests and values drive the process, but rather deferring to and facilitating their interests and values. In that way, I tried to make myself useful to people struggling to survive in a harsh environment" (Patton, 2004: 278 & 279).

It was also during his stint in that country as community development generalist and facilitator, that the importance of putting to the fore interests of those who are supposed to be beneficiaries of development efforts was impressed upon him by his experiences in a situation of dire need. This influence filtered to his approach to evaluation as reflected in the utilization-focused evaluation's emphasis on the primacy of intended users - from the beginning to the end. He admits that his "approach to evaluation grew out of those seminal community development experiences in Africa. From the very beginning, it was clear to me that I was not going to be the primary user of the evaluation findings. My niche would be facilitating use by others" (idem: 279). He further describes those experiences as 'one of the early and lasting influences on my perspective'.

The sociology literature introduced Patton to and developed his interest in diffusion of knowledge theory as propagated by Everett M. Rogers. Patton mentions the literature on diffusion of innovations as one of the major sources of ideas for utilization-focused evaluation. It was in this literature that he came across certain characteristics that are fundamental to the adoption and dissemination of new information, which are generally essential for dissemination and adoption of new ideas in other fields. Patton (2004) pinpoints three specialisations within sociology that shaped his evaluation perspective; namely: 'diffusion of innovation, sociological perspectives on power and conflict, and organizational sociology, which examines how people behave in institutions and organizations', behaviour which he claims can be easily observed in programmes which

are one of the primary arena of evaluation research. There is no doubt about the influence of this literature in moulding Patton's approach to evaluation as well as its efficacy to the then emergent field of study which was confronted with a major challenge of utility. The works on the diffusion of innovations were central in shaping Patton's thoughts on communication and dissemination of information because of the theory's emphasis on the 'nature and role of communication in facilitating further dissemination within local communities'. Rogers is said to have deconstructed the meaning of development from being a pillar of strength to a dominant paradigm in development and communication research. The dominant paradigm is reported to have overemphasised the significance of causal hypotheses to the neglect of socio-structural, political and institutional issues. Srinivas R. Melkote 2006 writes as follows about Rogers:

> "First, he redefined the meaning of development, moving away from the earlier technocratic, overly materialistic and deterministic models. Rogers defined the 'new' development as 'a widely participatory process of social change in a society, intended to bring about both social and material advancement (including greater equality, freedom, and other valued qualities) for the majority of the people through their gaining greater control over their environment" (2006: 116).

These sentiments of advancement of the disadvantaged, people's involvement in and control of their environment and prioritisation of local context have an echo in utilization-focused evaluation. According to Patton (1986), the book "Utilization-Focused Evaluation" is a product of a study conducted 'with colleagues and students of 20 federal health evaluations to assess how their findings had been used and to identify the factors that affected varying degrees of use'. It is enlightening to note that respondents were required to answer questions that affected use as propounded in the diffusion of innovations theory. Patton declares: "That study marked the beginning of the formulation of utilization-focused evaluation. We asked respondents to comment on how, if at all, each of 11 factors extracted from literature on diffusion of innovations and evaluation utilization had affected use their study" (Patton, 2004: 282).

There is a remarkable convergence between the roles and positions of what Rogers calls change agents in diffusion and evaluators. Firstly, they both serve in capacities of intermediaries between powerful stakeholders (governments, funders etc.) and those whom the services are intended to benefit. Secondly, like evaluators who are mostly

highly educated and equipped with technical skills, Rogers writes that change agents also 'usually possess a high degree of expertise regarding the innovations that are being diffused. Most change agency personnel may have a Ph.D. in agriculture or a M.D degree or possess other technical training' (idem: 368). These two factors put them in an unenviable situation of being culturally both a part of and removed from both the sponsors and the beneficiaries of services. An inclination towards either side may rouse suspicion and distrust from one quarter, thus denting their credibility, which is critical for their success.

Rogers identifies certain roles or qualities of change agents which are vital for diffusion, these can be extrapolated to evaluation because of their relevance to evaluators. They are:

To develop a need for change – this refers to awareness creation role of change agents in order to create an environment for a change in behaviour on the part of the client. There are few prospects of adoption of whatever an evaluator suggests if stakeholders are convinced that they are on the right path. *To establish an information exchange relationship* – like evaluators, change agents have to win the trust of clients by establishing rapport with them. Change agents have to appear trustworthy, competent and credible by empathising with the needs and problems of clients. This is important in evaluation research especially so in qualitative research where a researcher is also an instrument. *To diagnose problems* – it is the responsibility of the change agent to analyse the conditions in order to understand the sources of problems. The agent has to do so from the client's standpoint. This principle is principal in utilization-focused evaluation. *To create intent in the client* – this is related to various strategies a change agent has to employ to convince clients of the need to adopt innovations. A user-focused evaluator also has to have a toolbox of multiple skills to communicate effectively and to motivate stakeholders to adopt results of evaluation, both positive and negative. *To translate intent into action* – the principle is slightly removed from utilization-focused evaluation in means of achieving ends, but ends themselves are analogous. The objective is to 'influence clients' behaviour change in accordance with recommendations based on the clients' needs. The same purpose applies to evaluation in general, the difference being that change agents use the services of opinion leaders as means to accomplish their goals, while utilization-focused evaluators rely on stakeholder involvement as an instrument of instilling trust in the evaluator as well as

fostering ownership of evaluation. *To stabilize adoption and prevent discontinuance* – once the behaviour is adopted, change agents work on stabilizing it by buttressing messages about innovations to clients. Clients are given assistance with regard to issues that need clarification to reduce uncertainty and how to use innovations. Process use serves this purpose in utilization-focused evaluation, and more, the evaluator guides and monitors use. Lastly, *to achieve terminal relationship* – this relates to the development of 'self-renewing behaviour on the part of the client'. At some stage the change agent, like the evaluator has to move on, but they must empower clients to be able to implement what they learnt in future and different environments.

There is a discernible influence of these change agent roles on utilization-focused evaluation. There is congruence between the 'creation of intent in the client' and Patton's engendering of commitment to use. This commitment must be 'translated into action' in order for use to occur. Like change agents, utilization-focused evaluators have to 'stabilise adoption and discontinuance' through guidance during the processes, dissemination and follow-up visits to ensure use. Terminal relationship in utilization-focused evaluation is achieved by developing evaluative thinking in stakeholders through a participatory process.

2.3.3 Exposure to qualitative research and Egon Guba's reaffirmation

Patton states in "Evaluation Roots: Tracing Theorists' Views and Influences" (2004) that he was first introduced to the quantitative-qualitative paradigm debate during his two year research for a doctoral dissertation at the New School for Behavioral Studies in Education at the University of North Dakota in 1972-3. It was at that institution that Patton, who had hitherto been schooled in quantitative methods, was introduced to the qualitative paradigm under the tutelage of Vito Perrone. He professes that he was particularly attracted to Perrone's 'philosophy of open education that emphasized the importance and value of each individual', hence his dedication of the first edition of the book "Utilization-Focused Evaluation" to Perrone. The manner in which Patton 'learned the politics of methods' must have left an indelible mark on him because he effectively had to do what almost amounted to two dissertations to appease both sides of the then raging paradigms debate. On the one hand, he had to produce a qualitative evaluation for North Dakota users. On the other hand, his dissertation committee, in line with its quantitative inclination demanded a quantitative study in the form of a linear regression analysis. According to Patton, he had to use all his skills in coding to

convert qualitative data to permit regression analysis, an entirely statistical dissertation for his doctorate.

The monograph, "Alternative Evaluation Research Paradigm" (1975), is Patton's contribution to the then intense debate between quantitative and qualitative research paradigms. He asserts that the beginning of the second half of the twentieth century saw a surge in studies in the field of programme evaluation. The disconcerting feature of this literature was that it was overwhelmingly skewed in favour of one paradigm, the quantitative paradigm, sometimes referred to as the 'dominant paradigm'. This model, based on natural science model, was unquestioningly presented as the only credible route to 'cumulative scientific knowledge' by esteemed members of the evaluation community at that stage, he contends. The monograph can be construed as a reaction to the domination of the quantitative model over the qualitative one. No pretence of neutrality in weighing the strengths and weaknesses of both paradigms because that might compromise the purpose of the monograph; which is to wrench the alternative model from subordination by the scientific one. Patton explains: "At the outset I considered the possibility of attempting to describe and contrast the two paradigms in a neutral fashion. However, the very dominance of one paradigm, the natural science model, and the subordination of the second paradigm, the alternative paradigm, convinced me that it is more important to attack this imbalance than to maintain neutrality" (1975: 16). Patton presents a case for an alternative model, and for ease of 'description and analysis', he categorised the paradigms in sets of dichotomous polars, even though he concedes that they are not strict dichotomies. They are: qualitative versus quantitative methodologies, validity versus reliability, subjectivity versus objectivity, closeness to versus distance from data, holistic versus component analysis, process versus outcome evaluation and research for practitioners versus evaluation for scientists. Each dichotomy is afforded a chapter in the monograph, and in line with the attacking mode, the weaknesses of the scientific model and the strengths of the alternative paradigm are highlighted. In the end, the monograph is both a good case for the alternative model and an appeal for inclusiveness, taking into consideration the evaluation needs of each programme.

Patton recalls that one of the episodes that left an indelible impression on him is Egon Guba's keynote address on 'Overcoming Resistance to Evaluation' at the Second Conference of the North Dakota Study Group on Evaluation at the University North

Dakota in 1977. Patton writes that at that stage, educational evaluation was dominated by experimental and standardised testing with no credible alternatives. Therefore, the study group was composed of educators who sought to develop "innovative approaches to evaluation that were appropriate for and consistent with progressive education with its emphasis on individualized outcomes, experiential, child-centred learning in the tradition of John Dewey and multi-age classrooms with an integrated curriculum" (Patton, 2008: 318). Guba's impression on Patton was considerable not only because he was a brilliant scholar and methodologist from a prestigious 'Big Ten' university, but also because he was 'assertive, forceful and unapologetic' in discussing qualitative and collaborative approaches to evaluation. Guba declared:

> "it is my experience that sometimes evaluators adopted a very supercilious attitude with respect to their clients; their presumptuousness and arrogance are sometimes overwhelming. We treat the client as a "childlike" person who needs to be taken in hand; as an ignoramus who cannot possibly understand the tactics and strategies that we will bring to bear; as someone who doesn't appreciate the questions he ought to ask until we tell him – and what we tell him often reflects our own biases and interests rather than the problems with which the client is actually beset" (Guba [1978] as cited by Patton, 2008: 318-9).

Patton confesses that up to that day, he still harboured doubts about the legitimacy of more qualitative and collaborative approaches to evaluation, but after the presentation and the deliberations with Guba, he felt a new sense of affirmation. In the monograph published in 1978, Guba argues for naturalistic approach and he highlights instances whereby it may be a better alternative to the conventional experimental mode of research. They are outlined thus:

- To enlarge the arsenal of investigative strategies for dealing with emergent questions of interest – Guba points to the fact that paradigms, being borrowed from other disciplines are not ideally suited to deal with important social questions. He laments the then widespread tendency to fit research problems into existing models that are a result of model building and hypothesis testing that are practised in experimental approach. These designs also tend to strip research of its context because of different groups they create, thus leaving out important information that could be gleaned from it.

- To provide an acceptable basis for studying process – there is a suggestion, albeit oblique that the adoption of the "black box" concept from physics by social scientists who follow the experimental method is misguided. Guba writes:

- "It was proposed that physics be studied as though it took place in a tightly sealed and inaccessible "black box", which had various knobs, levers, wheels, pointers, and dials sticking out of it. The object of the game was to determine how any given manipulation of the knob, lever, or wheel might influence, say, a dial or pointer reading; and then to guess what must be going on inside the box from external evidence of these manipulations (Guba, 1978: 24).

- It is argued that physics adopted this strategy because internal processes were not accessible to direct observation. Therefore, it is preferable to observe processes directly where possible rather than to infer them through some manipulation.
- To provide an alternative where it is impossible to meet the technical assumptions of the experimental approach in the real world – The consistency of laboratory experiments when applied to the real world environment is questioned. It is asserted that experimental approach draws its power from control of different variables, but some of the assumptions from laboratory experiments are not applicable to the real world because of the absence of planned variations. What Patton refers to as the 'personal factor' is too strong in the real environment – human, political, social, economic and other factors conspire against them.
- To better access the implications of treatment–situation interaction – contention is made that higher-order interactions are very crucial in social research, particularly in evaluation research. Naturalistic inquiry, because of its emphasis on detailed descriptions of interactions, is said to be better placed in facilitating the understanding of such interactions. On the other hand, interaction effects are taken to be relatively insignificant in conventional inquiry.
- To redress the balance between reconstructed logic and logic-in-use – Attention is drawn to the distinction between reconstructed logic and logic-in-use. Abraham Kaplan (1964) as cited by Guba wrote:

- "We have physiologies and histories, and some of us also think and write about these things. Similarly, scientists and philosophers use a logic – they have a cognitive style which is more or less logical – and some of them also formulate it explicitly. I call the former *logic-in-use* and the latter the *reconstructed logic*. We can no more take them to be identical or even assume an exact correspondence between them, than we can in the case of the decline of Rome and Gibbon's account of it, a patient's fever and his physician's explanation of it (Guba, 1978: 27).

- This logic is used by Guba to argue that there is a difference between what inquirers do and reports of what they do. In the same vein, scientific methods are reconstructed logic with little connection with logic-in-use, which purportedly is the real thing. The problem with this line of argument is that it is so extreme that it annuls the whole communication enterprise. If what people say is not what they think, then we have a problem – what we read in this statement for naturalistic inquiry may as well not be what readers perceive it to be.
- To avoid implicit shaping of possible outcomes – It is averred that two elements shape the outcomes of inquiries – what is there to be studied and what the investigator introduces in the process of studying it. In experimental inquiry, there is *a priori* manipulation of certain variables which may artificially determine the scope and findings of an inquiry, thus giving a distorted picture of a given situation. The significance of this argument is that it draws attention to the danger of linking or varying certain variables, which results in the presence or absence of their influence in the findings. But in a naturalistic inquiry, all the variables are studied in their natural environment without any interference.
- To optimise generalisability – even though conventional inquiry prioritises generalisability, its requirement for external manipulation imposes some impediment to the very intention because the context outside the experiment cannot be produced like the one in the experiment. Moreover, generalisations do decay as attested to by "Bronfenbrenner's conclusion that class differences in parenting observed in the 1950s were just the reverse of those observed in 1930" (Guba, 1978: 29).

There are unmistakable traces of Guba's influence in the singular importance Patton attaches to situational responsiveness, that is, evaluations should be designed with careful attention to contextual factors and be responsive to them. Guba criticised the practice of fitting research problems to already existent designs prevalent in experimental approach. Situational responsiveness also means studies conducted on the basis of real world data as opposed to laboratory controlled environment that scientific studies encourage. It also means that there is less emphasis on generalisability since contexts are seldom similar socially, economically, culturally and politically. Interactions in the social realm seldom lend themselves to experiments, hence the need for qualitative studies in applied social sciences. Although Patton is not as strongly pro-qualitative paradigm as Guba is, there is no mistaking Patton's disposition for qualitative paradigm. The number of books he wrote in qualitative studies attests to that.

2.3.4 Influence of John Dewey's Progressive Education and Pragmatism

The seeds of progressive education as propagated by John Dewey, one of the advocates of pragmatism in education, are also evident in Patton's utilization-focused evaluation.

Patton wrote that he had been inspired by the experiential approach of the North Dakota Study Group to Evaluation which was consistent with John Dewey's progressive education, under the guidance of Vito Perrone. Contrary to traditional education on which the experimental paradigm with its standardized methods and teachers unloading their knowledge into students without understanding the latter's experiences was based; progressive education introduced learners for whom education is meant to the centre of the educational system. Dewey argued that each person is different both genetically and in terms of experiences. Therefore, curricula should be designed in such a manner that they accommodate individual differences and learner freedom. Consistent with his overall pragmatism, with its emphasis on the primacy of practice, Dewey intimated that education is not an end in itself. Rather, it has a broader social purpose of making people more effective members in a democratic society, hence the rejection of a top-down authoritarian system of pedagogy in traditional education. Pragmatism claims that the meaning of a proposition is to be found in the practical consequences and that impractical ideas are to be rejected. Nevertheless, Dewey understood that freedom of learners alone cannot solve educational problems and that it can be as problematic as authoritarianism in traditional education. He therefore advocated reconciliation of

traditional and progressive education by acknowledging the strengths and weaknesses of each approach as elucidated in "Experience and Education" (1938).

In this book John Dewey clarifies some issues about education that appear to have been central to optimum acquisition of knowledge by learners then and are still pertinent at present. The issue of concern is the exclusive dichotomous opposition between the 'traditional' and the 'progressive' approaches to education which does not allow space for cooption of positive input from both schools of thought to shape a holistic approach to education. The traditional approach perceives the role of education as preparation of learners 'for future responsibilities and for success in life'(1938: 2) by imparting in them the subject matter that is composed of skills, bodies of knowledge and rules of conduct which have already been developed from the past. Teachers and books serve as media through which these are conveyed as well as sources. Progressive educators deem the aims and methods of the former as an imposition on the learners hence their advocacy for free expression and individuality. They claim that the traditional education is based on misguided assumption that the finished product from the past that the learners are taught will be eternally relevant, which presupposes a static world. Dewey's full analysis of the two approaches places the opposition between the traditional and progressive education in terms of dichotomies thus:

> "To imposition from above is opposed expression and cultivation of individuality; to external discipline is opposed free activity; to learning from texts and teachers, learning through experience; to acquisition of isolated skills and techniques by drill, is opposed acquisition of them as a means of attaining ends which make direct vital appeal; to preparation for more or less remote future is opposed making the most of the opportunities of present life; to static aims and materials is opposed acquaintance with the changing world" (1938: 5-6).

While the new system may sound advanced, Dewey cautions against the negative philosophy whose point of departure is 'rejection' and 'self-aggrandisement'. This may lead the proponents of the new system to pay little or no attention to a rigorous examination of its own underlying principles, which are critical to its validity and strength. For instance, the wholesale rejection of outside authority – everything that teachers instruct learners cannot be of zero value – otherwise the very critics including Dewey would not be in a position to exercise such mental ability to have input in the appraisal of matters of such high intellectual calibre. He therefore, propagates the

middle ground whereby education in schools begins with what experience and interaction have already taught learners from their pre-school life and build cumulatively on those to produce a more mature and rounded person. However, he highlights the abstractness of principles when they are not applied to produce concrete consequences. He argues that "it is a sound educational principle that students should be introduced to scientific subject-matter and be initiated into its facts and laws through acquaintance with everyday social applications" (Dewey, 1916: 98). These principles are consonant with those of applied research or evaluation research particularly utilization-focused evaluation in its emphasis on use. For reason to be useful it has to be pragmatic. According to Dewey, education should have purposes for both the individual student and for society. As such, educators should provide learners with experiences that are immediately valuable to them and that enable them to contribute better to society. His progressive education was against the traditional practice in education for what he perceived as a lack of holistic understanding of students and its curricula, which emphasised content only, instead of stressing both content and process. For him, process is essential as it introduces the pragmatic dimension to education because education should not be only about the well-being of individual students, but there must be processes of transfer of the assimilated content into practice in a societal setting. It is that pragmatism which appealed to Patton and undergirds his logic in utilization-focused evaluation in so far as it acknowledges theory building as well as it emphasises use. Dewey's influence on Patton is particularly apparent in the latter's approach to paradigms debate in "Logical Incompatibilities and Pragmatism", when he writes:

> "My approach is *not* to try and resolve the paradigms debate. I try to establish an environment of tolerance and respect for different, competing viewpoints, and then focus the discussion on the actual information that is needed by each group: test scores? Interviews? Observations?….Multiple methods and multiple measures will give each group what they want" (Patton, 1985: 307).

2.4 Conclusion

Patton's ideas are a consequence of a combination of factors that range from academic training as well as experience acquired through fieldwork. Academic training was particularly useful in imparting quantitative methodological skills, beginning with tuition in quantitative methods early in his graduate studies at the University of

Wisconsin. Sociological literature introduced him to the works of Hage, Etzioni, and Crozier among others which played a pivotal role of grounding him in various aspects such as the philosophy of science, theory construction and so forth. The writings of Rogers on diffusion of innovations should not be underestimated in shaping Patton's perspective and approach to utilisation of evaluation processes and findings. It was the "Diffusion of Innovations" which instilled in Patton an interest of methods of knowledge dissemination to target populations, which lays the basis for use. However, the defining moment in his thinking was his encounter with the University of North Dakota Study Group on Evaluation which initiated him into qualitative paradigm. The study group introduced a counter-balance to his quantitative training by initiating him into qualitative learning in the field, which he pursued with passion. It was in that institution where he met Egon Guba who reaffirmed his commitment to qualitative methods. The conflict between the qualitative predilection of the study group on evaluation where Patton was doing fieldwork and the quantitative preference of the University of Wisconsin Sociology Department where he was doing his doctoral dissertation, meant that he was caught at the centre of the paradigms debate and had no choice but to adapt to his predicament. He eventually presented a qualitative report to the University of North Dakota Study Group on Evaluation and had to quantify qualitative data for his dissertation at the University of Wisconsin.

With regards to experience, Patton's stint in Burkina Faso as a Peace Corps was instrumental in teaching him to put those supposed to be beneficiaries of projects or programmes first, thus effectively registering in him the significance of reaching out to intended users and of personal involvement. That idea was further ingrained in him by the study he conducted in the mid-1970s to assess the use of findings in 20 federal health evaluations. All in all, the ideas which ultimately found their way to "Utilization-Focused Evaluation" are a product of an intersection of learning and some difficult experiences in the field.

CHAPTER THREE

PATTON'S NOTION OF UTILISATION

3.1 Introduction

This chapter is composed of the definitions of evaluation research, utility and use from Patton's perspective, a perspective which changed over time. The definitions are necessary to follow his writings. Positivism gave the false hope of social betterment. Patton's focus on use was not only a reaction against the experimental paradigm which bore little fruit as far as use was concerned, but also against positivism that promoted it. Positivism proposes that the logic of inquiry is supposed to be the same across all sciences, a proposition which caused resistance from proponents of the qualitative paradigm and consequently led to the paradigms debate in social sciences. Patton is opposed to the centrality of the scientific method in social development, particularly in applied sciences of which evaluation research is part. Therefore, his approach to evaluation is not only a reaction to scientific method, but also a rejection of positivism which actually enacted the scientific method.

The so-called utilisation crisis highlighted the meaninglessness of conducting evaluations employing predominantly experimental designs without paying due attention to utilisation. The fact that evaluation findings were little used or not used at all did not justify clients pumping in financial resources into a practice that did not bring returns on their investments. It led directly to focus on use in the sense that utilisation-focused evaluation. Several barriers to use both empirical and non-empirical approaches decreased evaluation use. The importance of use is manifested in the multiple types of use, beneficiaries and spheres of evaluation utilisation as propounded by Patton. Beneficiaries and spheres of use help to elucidate the importance of use by indicating the scope and levels of users. Misuse is treated separately from use.

3.2 The meaning (s) of "evaluation research"

We start our discussion with defining some key terms: "evaluation research", "use" and "utility". Patton's first definition of "evaluation research" appears in the 1978 edition of "Utilization-Focused Evaluation". It is noteworthy that his definition evolved three times from the 1978 edition of utilization-focused evaluation to its 1986 edition. These permutations are in part an appreciation of the difficulties attendant to the definition of evaluation and the fine-tuning it thereof. The first, definition of "evaluation research"

which was a working definition for selection of the cases from the Office of Health Evaluation for inclusion in the sample for the project on utilisation reads thus: "Evaluation research is the systematic collection of information about the activities and outcomes of actual programs in order for interested persons to make judgments about specific aspects of what the program is doing and affecting" (Patton, 1978: 26). This definition is crucial because of its focus on respondents' reactions to questions on utilisation imposed on them, which ultimately formed the basis of the notion of utilization-focused evaluation. One peculiar aspect of this definition is the strong summative intent of evaluation contained in it. There is no reference whatsoever to improvement in the purpose of evaluation, only judgement. In "Evaluation: A systematic Approach" (3rd ed), Rossi and Freeman define evaluation research in slightly different terms from Patton, but they approximate Patton's definition of programme evaluation in their inclusion of utility of social programmes. They write that "evaluation research is the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs" (Rossi and Freeman, 1985: 19). This definition suggests a compromise on some scientific principles to cater for utility while simultaneously ensuring that methods are rigorous enough to give credibility to the study.

The intricacy of definition becomes apparent when Patton discusses the lack of consensus among evaluators about the definition of evaluation, which has ramifications for evaluation use. Put differently, without agreement on evaluation definition, there can be no consensual discourse about its use. He illustrates this by pointing at the review of literature done by William J. Gephart on evaluation to find out evaluators' view of evaluation in "Six Honest Serving Men for Evaluation" (1985). Gephart discusses six different definitions, each related to the approach used by a certain section of evaluation researchers/practitioners:

> "(1) His classificatory definition of evaluation describes evaluation as a "problem-solving strategy" employed for establishing the relative or absolute worth of various choices. (2) His comparative definition likens evaluation to research, development, management, and other problem-solving strategies, pointing out similarities and differences of each. (3) His operational definition tells how an evaluation is conducted, from the identification of the impending decision through data collection and analysis to information use. (4) His componential definition explains that evaluations include a problem, a situation

> involving choices, data on the worth of the options, a context, a set of values, a time frame, and so on. (5) His ostensive definition gives examples of evaluations (e.g., deciding which dishwasher to buy). (6) His synonym definition includes such words as judgment and appraisal" (Gephart [1981] as cited by Patton, 1985: 2-3).

All these definitions emphasise what different evaluators deem as important in their practice and they have implications for design and use. In "Six Honest Serving Men for Evaluation**"** (1985), Patton attempts to cover as much scope as possible in his definition of evaluation while simultaneously avoiding unworkable breadth by harnessing the definition to focus on specific people. That can be defended against charges of parochialism because trying to please everybody can be counter-productive as it can lead to the dissatisfaction of everybody. He defines it in these terms:

> "The practice of evaluation involves the systematic collection of information about activities, characteristics and outcomes of programs, personnel, and products for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs, personnel or products are doing and affecting. This definition of evaluation emphasizes (1) the systematic collection of information about (2) a broad range of topics (3) for use by specific people (4) for a variety of purposes" (Patton, 1985: 5).

This definition is certainly more elaborate than the previous one, with the addition of personnel and products in its coverage. There is mention of improvement this time around, but there is toning down of expectations from evaluations in the sense that people must not expect more than a reduction of uncertainty. Nonetheless, this is a bold step from someone who is aware of the intricacies involved. That might be the reason for a few, albeit telling changes in the 1986 edition. In this version, the 'practice of evaluation' had been replaced by 'program evaluation', which is definitely a narrowing down to evaluation of programmes only. This is corroborated by the exclusion of 'personnel' and 'products' from the range of disciplines covered by the definition. It is also important to note that a distinction is drawn between *programme evaluation* and *evaluation research*. Patton asserts by way of clarification that: "This book emphasizes utility, relevance, practicality and meeting the information needs of specific decision makers, thus the focus of program *evaluation*. The term "evaluation *research*", will be used to refer to studies of program outcomes and effects where there is relatively greater emphasis on generalisability, causality, and credibility within the research

community" (idem: 15). There is a suggestion that the latter is less useful, almost to a degree of uselessness to community development in the oppositional language employed to shed light on the differences. While programme evaluation has specific users quite apart from the benefits accruing to evaluation researchers, in the form of decision makers, evaluation research appears to be limited to the research community. This effectively categorises programme evaluation and evaluation research into what Shadish and Epstein called *service* and *academic* evaluators in "Patterns of Program Evaluation Practice among Members of The evaluation Research Society and Evaluation Network" (1987). They write that the "general distinction found in the present study between academic and service evaluation had been predicted in general terms by several authors (Cook and Shadish, 1986; Cordray and Lipsey, 1987). Although the patterns found in our study do not perfectly support patterns that have been proposed in the literature, the general discrepancy between service-oriented and academically oriented evaluators now seems warranted on both theoretical and empirical grounds" (Shadish and Epstein, 1987: 587). They went on to suggest that the variations between them are attributable to 'different demands and reward structures under which the two groups often operate'. It is intriguing to see the development of Patton's definition from a synthesis of programme evaluation and evaluation research to a stage where the two are almost diametrically opposed.

There was a general consensus among scholars of evaluation and prominent reviewers in the 1960's and 1970's that evaluation research had inconsequential impact on decision making, but they did not provide any operational definition of utilisation. In the absence of a clear definition of utilisation, Patton works out what it is from definitions found in the literature. He concludes that the commonly held view of utilisation is that it "occurs when there is an immediate, concrete, and observable effect on specific decisions and program activities resulting directly from evaluation research findings" (Patton, 1978: 24). But after the findings of the study on evaluation utilisation conducted in the Office of Health Evaluation, Patton and his colleagues discovered that the picture of use was not as gloomy as reviewers painted it. The contention is that there was in fact utilisation, but it occurred in ways not anticipated, that is, 'immediate, concrete and observable effects' on decision making. There is a suggestion that this view of utilisation might have contributed towards negative sentiments against evaluation research and even acted as a constraint to it.

The findings of their study indicated that utilisation is dependent on the values of users, hence Patton's drive for inclusion of intended users in all phases of evaluation. His alternative definition reads thus: "Reducing uncertainty, speeding things up, and getting things started are the real impacts – not revolutionary, organization-shaking impacts – but real, important impacts in the opinion of the people we interviewed" (1978: 31). This is perhaps indicative of the complexity of the subject. Patton, on the other hand, defines utilisation in "The Evaluators Responsibility for Utilization" thus: "By utilization I mean *intended use by intended users*" (1988a: 14). This definition relegates the definition of use to intended users whose view it is that utilisation has occurred. This widespread perception of utilisation definition played a role in depicting use of evaluation as negligible relative to multitudes of dollars invested in evaluation, and therefore focused attention on it.

Utility undergirds Patton's approach in utilization-focused evaluation and indeed, it appears to be the central theme running through the whole of his works on evaluation. In "Use as a Criterion of Quality in Evaluation", he writes that:

> "Utilization-Focused Evaluation begins with the premise that evaluations should be judged by their 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. Nor is use an abstraction. Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the *focus* in utilization-focused evaluation is on *intended use by intended users*" (2001: 160).

However, it is difficult to pin down any single definitive meaning of utility because it is used so broadly that it almost runs across all the 1981 Joint Committee's Standards for Excellence in Evaluation. The 1986 edition of utilization-focused evaluation refers to "utility test" (whether data are relevant) (1986: 222), "useful and relevant information" (idem: 235), "utility tests" (whether data are useful) (idem: 243), "The Utility Standards are intended to ensure that an evaluation will serve the *practical* information needs of given audiences" (idem: 26 emphasis mine). It is therefore, not far-fetched to submit that Patton's idea of utility of evaluation emphasises the usefulness of evaluation. In order for it to be useful, it has to be relevant to intended users as well as practical. Feasibility is one of the standards. This understanding is further entrenched in "Towards Utility in Reviews of Multivocal Literatures" (1991), where Patton

pronounces on utility along these lines: "When I undertake a program evaluation or policy analysis, I am doing so to provide *useful* information to specific decision makers and stakeholders (people with a vested interest in the program and, therefore, the evaluation). They want information they can *use*. The information must be *relevant*, *understandable*, and *accurate*. It also usually has to be generated in a hurry. When a multivocal literature review is part of the process, its purpose is to find out what others are saying, writing and doing that may be *useful* to the evaluation's intended users" (Patton, 1991: 287 emphasis mine). It is significant to note the inclusion of accuracy in addition to relevance and comprehensibility in Patton's discussion. Accuracy used to be inseparable from rigour in the traditional scientific paradigm and indeed, methodological rigour and technical quality were used as *the* criteria for judging evaluations prior to the paradigms debate rage of the 1970s. The question of utility is complex because it is intertwined with use, credibility, validity, quality and so forth – to be touched on later. At this point, it will suffice to emphasise that Patton rates utility ahead of all other standards of evaluation excellence (utility, feasibility, propriety and accuracy – in that order) and one can't help detecting a note of triumph when he quotes the Chair of the Committee, The Joint Committee on Evaluation Standards, Dan Stufflebeam when he wrote: "The standards that will be published essentially call for evaluations that have four features. These are *utility*, *feasibility*, *propriety* and *accuracy*. And I think it is interesting that the Joint Committee decided on that particular order. Their rationale is that an evaluation should not be done at all if there is no prospect of its being useful to some audience." (Stufflebeam [1980] as cited by Patton, 1986: 25). He joyfully declares that the order is not necessarily a denigration of rigour, but that "Rigor is simply made the servant of utility" (Patton, 1991: 288). From Patton's perspective, utility is strongly tied to the capacity to be used by intended users (usefulness) in line with his strong conviction that usefulness of an evaluation is determined by those who use it. Utility, it seems, is the precondition for conducting evaluations; being useful means realising its intended purpose to intended users through any type of use.

3.3 Reasons for Patton's focus on use

3.3.1 Unmet expectations created by positivism, modern science and rationality

Patton describes the utilisation problem by alluding to the implicit expectations of science by government. "In my judgement, the utilization of evaluation research represents such a crisis in institutional arrangements. This issue of utilization has

emerged at the interface between science and government" (Patton, 1978: 12). There was apparently, a perception that modern science could be employed to rid society of its ills because of its perceived reliance on empirical data. The logic of rationality presumed that science (evaluation research) would provide reliable data on what worked and what did not work; what worked would then be taken up and used for planning by politicians. These plans were expected to be implemented to the letter by dedicated development workers, thus, culminating in social upliftment.

These sentiments related to the promise of modern science as a panacea for social ills resonate with key elements in Jürgen Habermas' theory of knowledge and its three cognitive interests that purportedly govern humanity's search for knowledge. The most pertinent interest in this instance is the emancipatory interest, of which it is written: "This is an interest in reason, in the human capacity to be self-reflective and self-determining, to act rationally. As a result of it, knowledge is generated which enhances autonomy and responsibility: hence it is an emancipatory interest" (Babbie and Mouton, 2001: 34). It was this promise of science for human emancipation and social progress and assumptions about human rationality and concomitant responsibility in the creation of a better world which irked different stakeholders when evaluation was not utilised the way they anticipated because they had a certain view of science, as a solution to social problems. This led to a so-called "utilisation crisis".

3.3.2 The utilisation crisis

The "utilisation crisis" can be understood against the backdrop of billions of dollars which successive governments of United States of America poured into evaluation research in its campaign to eliminate social problems such as poverty, unemployment, diseases and other social ills. The period following the Second World War saw competition between the Western Bloc led by USA and the Eastern Bloc under the leadership of Russia as the controller of the then Union of Soviet Socialist Republics. That was aggravated by the launch of Sputnik to space by the Russians. The USA then embarked on a mass education campaign because they felt that they were lagging behind in that area and they consequently invested huge sums of money in educational problems and others aimed at getting rid of all social ills under the banner of the War on Poverty and the Great Society Programmes. By the 1970s, in the wake of the Vietnam War, conditions were not ideal for a continued splash of money for programmes, therefore, governments sought means to document proof of the efficacy of

programmes in fighting poverty - hence the recourse to evaluation research. Evaluation research gained ascendancy as the US government's mechanism of ensuring accountability by those entrusted with the task of improving human conditions. However, evaluation results proved to be less fruitful than stakeholders expected; accusations and counter accusations were hurled around. On the one hand, evaluators blamed decision makers for not using research findings and on the other hand, decision makers retorted that findings did not respond to their problems. It was this crisis of utilisation of evaluation findings that ultimately prodded Patton and his colleagues to embark on a study of evaluation use which became the basis of utilisation-focused evaluation. He writes: "Then the advisory board chairperson turns to me in exasperation and asks: "Tell us, what do we have to do to produce good evaluation research that actually gets used?"" (1986: 12).

3.4 Barriers to use

This section highlights some of the impediments to the conduct of empirical evaluations and pitfalls associated with evaluators. These contribute to the production of poor evaluations and may have contributed in part to nonutilisation of certain studies because of their low quality, thus exacerbating the utilisation crisis, Patton suggests. The first two barriers are examples of non-empirical barriers to use and the others are factors inhibiting use as a result of indiscretions on the part of evaluators despite their quest for empirical data.

- **Charity orientation –** Patton brings to our attention some of the practices that constrained utility of programmes in the days preceding evaluation research. Prior to the rise of evaluation research, programmes were evaluated on the basis of faithfulness of staff irrespective of their outcomes - he brands that a charity model. He states that in this model "the criterion for evaluation of programs is the sincerity of funders and program staff; the measure of program worth is that program organizers care enough to try their best to help the less fortunate" (Patton, 1978: 13). While it is crucial to certify the nobility of intentions on the part of funders to ensure that they are aimed at improvement of human conditions and that staff perform their duties meticulously, it is not a foregone conclusion that good intentions automatically result in good implementation of programmes. Sincerity and caring are crucial ingredients to the success of any programme, but that is more on the side of management. The focus of

evaluation is to find out whether or not that sincerity and caring are channelled towards effective implementation of programmes as per goals and objectives and that findings are disseminated in time and in accessible form to be utilised.

- **The Pork-Barrel approach –** This approach is more or less similar to the symbolic use of evaluation results in that evaluation becomes an instrument of political manoeuvring. Patton puts it thus: "The second major model of evaluating programs has been the pork barrel approach, which takes as its outcome criteria for program evaluation the strength and leverage of the program's constituency" (1978: 14). In such instances, evaluation results are tailored to please or appease the politically dominant section at whose whims the programme may be conceived or be terminated, and evaluators become pawns in political games. Weiss had this to say about such practitioners: "Social scientists who not only address the research problems that officials set but who address them in the terms and with the value orientations of the incumbent officeholders abdicate their roles as scientists and become technicians for the powers-that-be" (Weiss, 1977: 2). Therefore, the pork barrel approach turned evaluators into technicians at the service of authorities at the expense of their personal reputation and to the detriment of the evaluation discipline in general.

- **Evaluator dominance –** Domination of an evaluation process occurs when an evaluator becomes a central player in an evaluation. At times an evaluator can find herself/himself in a situation where stakeholders are not sufficiently responsive to the evaluation process to the extent that s/he has to make decisions by herself/himself. The situation can be a consequence of a lack of necessary knowledge on the part of other stakeholders who constitute an evaluation task group and an evaluator is compelled to make decisions, thus becoming a decision maker by default. Intimidation, both intended and unintended (fear of the status of an evaluator), and the failure of an evaluator to facilitate an evaluation in such a manner that it encourages interaction or to select apposite primary stakeholders can result in evaluator dominance. Ultimately, the evaluation enterprise becomes an evaluation by the evaluator for the evaluator in a more or less similar fashion to scientific research with little prospects for use. This makes it imperative for the evaluator to select knowledgeable, assertive and dedicated primary intended users.

- **Identification with the audience and not the targeted beneficiaries –** Patton claims that 'identification of audience' can hamper full potential for utilisation because the audience is not necessarily primary users of evaluation organised for regular participation in the evaluation. He writes: "Audience refers to groups of largely anonymous faces: the "feds", state officials, the legislature, funders, clients, the program staff, the public, and so forth. But specific individuals are not identified from these audiences and organized in a manner that permits meaningful involvement in the evaluation process, the utilization is reduced" (Patton, 1986: 52). When consideration is given to the fact that utilisation does not come at the end of evaluation, but prepared for throughout the evaluation process, it is vital that specific people be identified to carry through the process to the end.

- **Decision-oriented approach –** The thrust of this approach is in decision and information to the neglect of decision makers and information users. "The decision-oriented approach stems from the rational social scientific model of how decision making occurs; that is, there is a clear-cut decision to be made; information is needed to make the decision; the social scientist supplies the information; and the decision is made with accordance to that information" (1986: 53). But the utilisation crisis has demonstrated that this model can at times be misguided. Certain evaluations have multiple stakeholders with various and sometimes competing interests. Evaluation information seldom serves all of them, implying that some will not be served. Furthermore, decision makers do not depend solely on evaluation; there are various sources of information on which decisions may be based. There is a need for identifiable people to make decisions, to use information and more importantly, to participate in all phases of an evaluation, it is rather presumptuous to think that decision makers will abandon other sources of information and rely on evaluation recommendations only.

- **Failure to identify primary information users –** Evaluators are cautioned against the assumption that evaluation funders are relevant evaluation information users by virtue of being funders. There are various reasons for supporting the conduct of an evaluation, least which, are for political use. An

evaluation may be commissioned to legitimise a programme, whereby the mere presence of evaluative activity, even if it does not have any positive contribution to the programme, can enhance its standing. Therefore, failure to identify primary intended users can reduce evaluation use.

- **Targeting organisations –** Potential for utilisation is reduced when organisations are identified as targets of evaluations. "When the organization is the identified consumer of evaluation, specific decision makers and information users do not have to be identified and organized, and usually they are not" (idem: 54). The contention is that individual people, not organisations, use evaluation information, therefore, targeting evaluations at organisations is like targeting them at faceless people or not targeting it at all.

One learns from Patton's discussion of these barriers not only the need to conduct evaluations which are data-based but also to exercise caution to remain focused on the goals of evaluation. Particular consideration should be given to identifying correct stakeholders or intended users so that evaluation attends to the needs of pertinent groups. Failure to do this may result in an evaluator becoming the centre of evaluation with a subsequent reduced relevance of evaluation or outright impertinence. Other factors, such as intimidation by powerful interest groups and inexperience may also lead to distortions or misevaluation. The common denominator in all these barriers is the evaluator's incompetence manifested through either agreeing to conduct non-empirical evaluations or misevaluations and less useful evaluations because of inexperience.

3.5 The importance of use

According to Patton, the importance of use is not addressed in a clear manner in the literature of utilisation. It is therefore put forward here that the significance of use is implicit in the manifold definitions of evaluation, its typology, purposes, beneficiaries and spheres of evaluation utilisation. These are addressed below:

3.5.1 The typology of utilisation

The title of the article, "Six Honest Serving Men for Evaluation" (1985), is adapted from Rudyard Kipling's poem Six Honest Serving Men. Patton perceives the questions

posed by the writer in the poem as relevant as a framework for organising his typology of evaluation. Rudyard Kipling's Six Honest Serving Men reads:

I keep six honest serving men.
They taught me all I knew:
Their names are What and Why and When
And How and Where and Who.

The interrogatives what, why, when, how, where and who are employed to provide a comprehensive typology of use. This paper builds on the groundwork begun in the first edition of utilization-focused evaluation and provides an example of its real world application in the Caribbean Agricultural Extension Project (1983).

A distinction is made between two types of impacts, the *conceptual* and *action* impacts. "Conceptual impacts are those which affect thinking about a program. Such uses may lead to conceptualizing implementation or outcomes in new ways, understanding dynamics of the program more thoroughly, or shifts in program priorities. Action impacts are those which lead to observable changes in actual operations of a program" (Patton, 1985: 7). Action impacts are what is commonly referred to as instrumental use and it is most noteworthy when there are visible changes in programmes as a result of evaluations. Such changes also affect decisions about programmes, whether to continue or to stop them, to fund them or not and a range of things over which decision makers have control. Patton points to a predisposition towards action impacts as indicated by the despair in the evaluation community about lack of use. However, he unequivocally dismisses the issue of hierarchy between the two when he says that "there can be no absolute standard which values action over thinking, changes in a program over keeping things the same, or decisions to do something over decisions to wait. There simply can be no hierarchy of impacts because the hierarchy is necessarily situational and depends on the values and the needs of the people for whom the evaluation is conducted" (idem: 8). Conceptual impacts are by their definition long term but they are just as useful as instrumental impacts, but in a different manner because changes in thinking are more enduring and affect the whole practice of evaluation once they are adopted and implemented, they produce among others, instrumental impacts. These types of use are direct uses, so is *process* use.

The latest edition of "Utilization-focused evaluation" (2008) refers to additional uses that have come into currency with the expansion of research on evaluation utilisation, among them is *influence*. Weiss highlighted the role of influence in her response to Patton to the effect that evaluators should aspire to influence decision making at a further distance from the programme being evaluated rather than acting as 'philosopher-kings' whose ideas would directly feed into decision making processes. But it was Kirkhart who defined influence in "Reconceptualizing Evaluation use: An Integrated Theory of Influence". She defined it thus: "the power of persons or things to produce effects on others by intangible or indirect means" (Kirkhart, 2000: 7). Enlightenment is separated from conceptual use but treated as its extension both in terms of compass and time. Weiss (as cited by Patton, 2008) wrote that: "*Enlightenment* adds a longer time dimension and connotes a broader policy scope to conceptual use. It involves the gradual percolation of ideas from evaluation into policy discourse, changing understandings, questioning assumptions, evoking new ideas, and altering practices"(Weiss [2004] as cited by Patton, 2008: 112). *Unintended effects* are use of unplanned, unpredictable evaluation processes or findings.

There is a temporal dimension to utilisation. Patton asserts that in the early days of evaluation attention was on *immediate* instrumental use. However, there was realisation that decision making processes are not always instant but take time; therefore, utilisation also was apt to be *incremental* from the time evaluation findings are released and beyond. "There are not a great many clear, specific and immediate decisions taken in public organizations. Rather, decision making tends to be a process of moving in a particular direction that is not always explicit and does not come from decisive moments of action" (Patton, 1985: 11). There is a problem however, in determining how long these incremental uses take. The longer they take, the more problems there are in attributing cause to any particular evaluation (measurement problem). Patton believes that there is still inclination towards immediate concrete and short-term uses even though he considers incremental use more important. This may be precisely because of the problem mentioned. The processes of evaluation are not discussed; they are imperative as they have both critical immediate and long-term uses.

Intentionality is reflected in certain kinds of use. For instance, evaluation use can be *planned*, that is, when intended users of an evaluation are identified at the onset and its use is judged by the extent to which it accomplished its intended or planned use. This is

relatively straight forward because the evaluator already has identified intended constituencies (intended users) from which to pick suitable candidates for the task team from. This is a case of intended use for intended users. "From this perspective, the most important question in researching the utilization of evaluation is whether the evaluation had its intended use. This solves the problem of defining utilization, addresses the question of who the evaluator is for, and builds in a time frame since the predetermined use would necessarily have a time frame" (idem: 12). Planned utilisation is appropriate for summative evaluation with its emphasis on accountability because the evaluator simply looks for feedback among intended users regarding services rendered. The shortcoming of planned use is that it is limited to those stated expectations and might miss other important long-term, unintended uses which are covered by unplanned use. *Unplanned* use is when an evaluation is carried out without any consideration to utilisation until after data analysis. This type of use is congruent with Scriven's notion of goal free evaluation.

Patton declares that "the debate that led to Scriven's proposal for goal free evaluation included concerns about attention to unanticipated consequences, side effects, and unstated goals as important outcomes of programs"(1985: 13). The formal/informal dimension follows the pattern where one form becomes the immediate centre of attention, only to realise that the other is just as important. *Formal* use revolves around the use of evaluation information that is neatly packaged in evaluation reports and publications and is usually observable to the public. This is the form that was used in the early evaluation days. However, the wisdom amassed through the years of practice alerted evaluators to the importance of *informal* use. "We have since learned that informal uses are often more important. This is the transfer of findings by word of mouth, in unplanned discussion groups, and in one-to-one interactions between the evaluator and program staff, administrators, and/or funders" (idem: 13). This is the use of the evaluator outside the formal processes of evaluation and may even go beyond it, the atmosphere is more relaxed allowing more varied inquiries, hence the allusion to its greater import than formal one.

Process use is perhaps the biggest discovery in the typology of use and an acknowledgement by Patton that he and other evaluators had overlooked a vital aspect of evaluation use by concentrating their energies on findings and their use (or misuse and abuse). It was only after two decades of doing evaluations that he came to

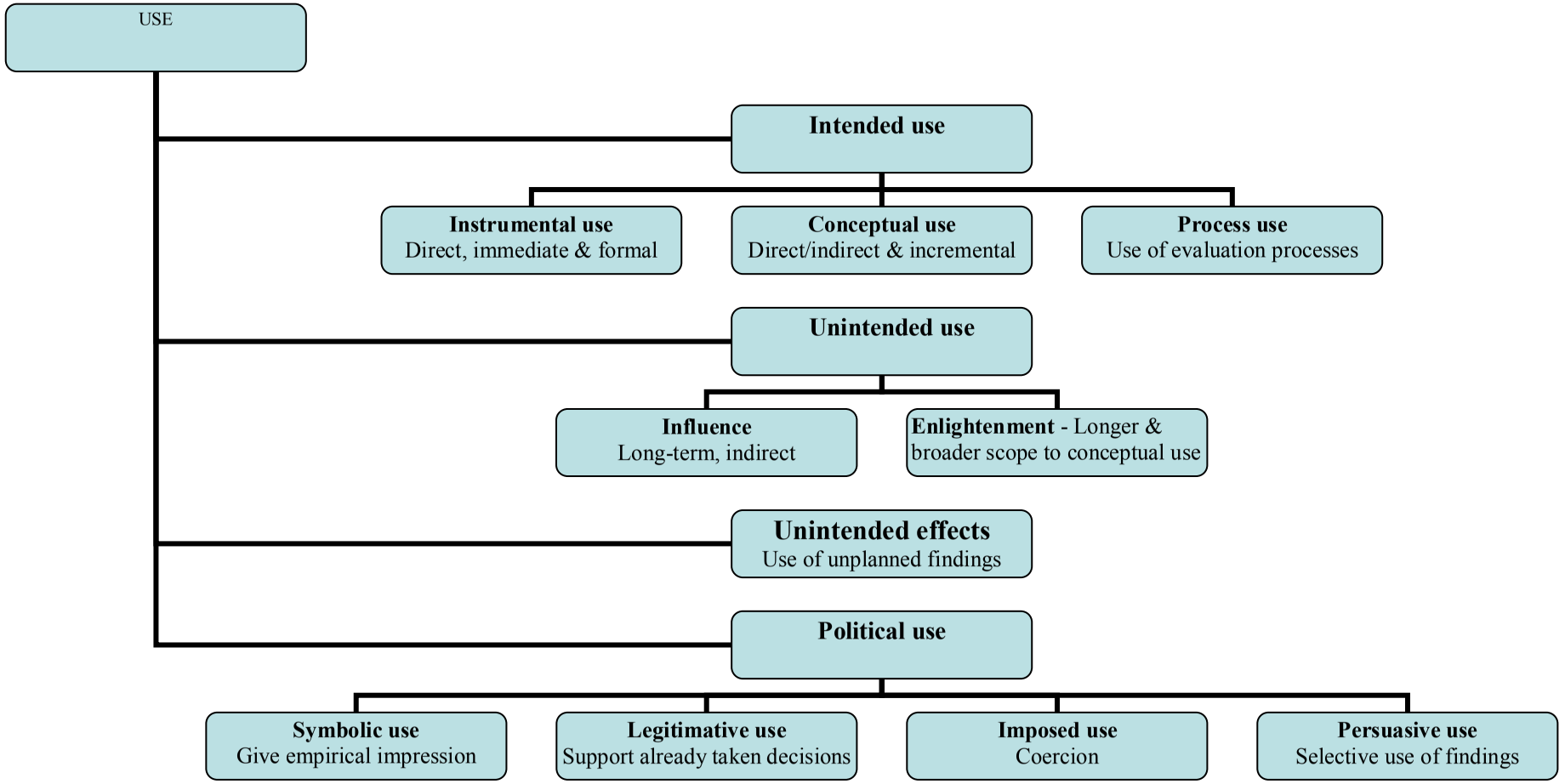
understand what many intended users meant in their responses to his follow up questions. He writes in "Discovering Process Use": "But beyond the focus on findings and recommendations, what they almost inevitably added was something to the effect that 'it wasn't the findings that were so important in the end, it was going through the process'" (Patton, 1998: 225). This he called process use and it is defined as "individual changes in thinking and behaving that occur among those in evaluation as a result of the learning that occurs during the evaluation process" (idem: 225). To elaborate on the point he draws attention to the fact that the evaluation field has its culture and with its values, ways of thinking and language; thus evaluation process is a 'cross-cultural experience'. The benefits of process use are critical because they do not only impart 'substantive knowledge' about something, but they teach intended users to 'learn how to learn', what Patton refers to as 'learning to think evaluatively'. The capacity to think evaluatively is more enduring in value than a set of findings. He declares:

> "Findings have a very short 'half-life' – to use a physical science metaphor. They deteriorate very quickly as the world changes rapidly. Specific findings typically have a small window of relevance… In contrast, learning to think evaluatively can have an ongoing impact. The experience of being involved in an evaluation then, for those stakeholders actually involved, can have a lasting impact on how they think, on their openness to reality testing, and on how they view the things they do. This is one kind of process use – learning to think evaluatively" (idem: 226 & 227).

Learning to think evaluatively capacitates staff and therefore contributes towards the overall development of an organisation. Another benefit of process use is to *enhance shared understanding* within programmes. The process of engagement by different stakeholders in the activities of an evaluation from goals clarification to data analysis engenders a sense of community, deepens communications among them which ultimately leads to an understanding of one another's points of departure because of improved comprehension of their backgrounds. Social, cultural and political groundings are critical in shaping one's perspective. Changes in programmes can occur as a result of this new found knowledge. The third kind of process use is that of an *evaluation as an intervention*. This is when the very process of evaluation sensitises stakeholders about certain aspects of their programme. For instance, interview questions can serve to direct attention to some neglected features of a programme. In this manner, going through the process serves to promote corrective action well before the findings are

derived, thus reinforcing a programme. *Instrumentation effects* help to focus attention of the programme on important issues because variables under scrutiny usually receive more attention than others. "The mantra of performance measurement – 'What gets measured gets done' – encapsulates one aspect of evaluation's process impact. What we choose to measure has an impact on how people behave. If staff or programs, for example, get rewarded (or punished) for those things that are measured, then those things take on added importance" (idem: 231). The potential downside can manifest itself in attention being directed exclusively at those aspects that are measured to the neglect of others, especially when stakes are high. This defeats the very purpose of evaluation. Two more types are added in "Process use as Usefulism" (2007). Fifth, *Infusing evaluative thinking into organisational culture* concerns the entrenchment of evaluation into an organisation so that it becomes part of the manner in which it does business. Although related to teaching stakeholders to learn to think evaluatively, it differs from it in that the latter is learning to learn at individual stakeholder level. It seems to be a step towards infusing evaluative thinking into an organisational culture which would be completed when evaluation has been deeply ensconced through a thoroughly functional internal evaluation unit within the organisation. The last type, *increasing participant engagement, self-determination and a sense of ownership* relates to the empowering nature associated with participation in deconstructing the highly technical evaluation jargon. Evaluation expertise and critical thought cease to be the sole preserve of evaluators, decentralising power to create multiple centres and a new sense of ownership of evaluation among stakeholders. It should be noted however, that each type has its downside.

Figure 3.1 Typology of use



3.5.2 Purposes of evaluation

This section addresses the different purposes for which evaluations are carried out. Evaluation is a strenuous activity both in terms of physical and mental application needed for its proper conduct. It is also demanding in temporal, spatial and financial resources. It is therefore apparent that there are imperative and valid reasons for doing evaluations. They include finding answers to questions related to effectiveness of programmes, policies or interventions in general, benefits of interventions relative to costs incurred, continuation or termination of programmes, improvements to programmes or of human conditions, accountability and so forth. The purposes are classified under certain broad categories depending on their main thrust. These purposes follow below.

3.5.2.1 Formative evaluation

Patton asserts that "formative evaluation typically connotes collecting data for a specific period of time, usually during start-up or pilot phase of a project, to improve implementation, solve unanticipated problems and make sure that participants are progressing toward desired outcomes" (2008: 118). He then stresses the Scriven factor in inventing the distinction between formative and summative evaluation. His definition is not different from Scriven's; it is only a matter of detail. According to Michael Scriven, "Formative evaluation is designed, done, and intended to support the process of improvement, and normally commissioned or done by, and delivered to, someone who can make improvements" (Scriven, 1991: 20). It is apparent from this definition that formative evaluation is aimed at improving programmes. This is usually accomplished by concentrating on finding causes for programmes' failure in order to redress them or strengths to build on. It effectively looks at the processes of evaluation to check whether it was designed in accordance with the goals of the programme and needs of stakeholders (intended users in the case of utilization-focused evaluation) and how it was implemented, that is, whether or not an evaluation has been implemented as it was designed. It also includes monitoring to a certain degree because these activities cannot be carried out without a monitoring instrument to determine shortcomings in design and implementation and other constraints. Although the stage at which formative evaluation can be done is subject to debate, it usually occurs in the early phases up to implementation. As an improvement-oriented evaluation approach, it is necessary that it be done in an ongoing way so as to discover any discrepancies between the expected direction and interim outputs of the programme in reality.

Formative evaluation assesses strengths to build on and weaknesses to discard in programmes for their better implementation. Formative evaluation activities include the collection and analysis of data during the programme and timely feedback of evaluation findings to programme staff to inform ongoing decision making and actions. Feedback is an important part of formative evaluation as it helps in fine-tuning the implementation of the programme.

3.5.2.2 Summative evaluation

Patton (2008: 114) asserts that "summative evaluation judges the overall effectiveness of a programme and is particularly important in making decisions about continuing or terminating an experimental programme or demonstration project. As such, summative evaluations are often requested by funders. …Summative evaluations provide data to support a judgement about the program's worth so that a decision can be made about the merit of continuing the program". For him, such evaluations are aimed at determining merit or worth are judgement-oriented, and therefore, summative evaluations are judgement-oriented. He draws a distinction between merit and worth with the former referring to the intrinsic value of a programme in meeting the needs of those it is intended to help. Worth, on the other hand, refers to the extrinsic value, that is, its value to those outside the programme, the larger community (idem: 113). For judgements to be made there should be explicit criteria upon which they are founded.

Again, this definition is not that different from Scriven's except that Scriven's definition incorporates everything other than formative, thus placing all evaluation purposes under the formative-summative dichotomy. "Summative evaluation is the rest of evaluation: in terms of intentions, it is evaluation done for, or by, any observers or decision makers (by contrast with developers) who need conclusions for any reasons besides developmental (Scriven, 1981: 20). Summative evaluation is therefore, judgement-oriented and it seeks to establish merit (intrinsic/inner value) or worth (extrinsic/contextual value) of a programme, policy or product being evaluated. The focus is on the effectiveness and efficiency of entities being evaluated, for example, whether or not they have accomplished what they were intended to; if so, was it in a cost-effective manner with optimal resources and so forth. At the end of a summative evaluation, a decision has to be made regarding its performance and to the effect of its continuance or termination. It follows then that the appropriate time to conduct summative evaluation is when the programme has been implemented or when there is

reason to suggest that it has been underway for quite some time or mature enough to expect it to yield some results. It is worth mentioning that Patton and Scriven have found common ground on this distinction. Patton has changed to assert that the distinction between formative and summative evaluations in one of roles. This is concurrence with Scriven's clarification that the differentiation is contextual, not of intrinsic nature.

3.5.2.3 Knowledge generation

An additional purpose of evaluation according to Patton is knowledge generation. He argues in "A World Larger than Formative and Summative" (1996) that while formative (improvement-oriented) and summative (judgement-oriented) evaluations are instrumental in use, conceptual use generates knowledge by influencing thinking on certain issues.

> "The evaluation *research* findings contribute by increasing knowledge. This knowledge can be as specific as clarifying a program's model, testing theory, distinguishing types of interventions, figuring out how to measure outcomes, generating lessons learned, and/or elaborating policy options. In other cases conceptual use is more vague, with users seeking to better understand the program; the findings, then, may reduce uncertainty, offer illumination, enlighten funders and staff about what participants really experience, enhance communications, and facilitate sharing of perceptions" (Patton, 1996: 132).

Among examples cited as knowledge generating/enlightening is *theory-driven evaluation* for its informative insights into general characteristics of successful programmes. According to Chen (1996), theory-driven evaluations can provide models for summative judgements or for improvement. The connection with social science theory also offers the potential for increasing knowledge about how effective programmes work, thus laying the basis for understanding what produces good results. This knowledge is not limited to particular programmes but is generally instrumental as guidance for future considerations for programme development and implementation. The same can be said of synthesis evaluations, which synthesise information from different studies to come up with a general pattern for a certain given variable under study. These syntheses represent an accumulated wisdom on particular aspects of evaluation that can be adapted for future use. For example, James Burry's synthesis of factors affecting utilization is a product of an extensive study of those aspects across

multiple evaluations and their utility. Conclusions drawn from such syntheses can be used to inform other evaluation studies. They are like case studies of lessons learned across many studies. They provide a wealth of information on processes and strategies necessary for successful implementation of those aspects synthesised. Patton writes:

> "Such generalizable evaluation findings about principles of effective programming have become the knowledge basis of our profession. Being knowledgeable about patterns of program effectiveness allows evaluators to provide guidance about development of new initiatives, policies, and strategies for implementation. Such contributions constitute the conceptual use of evaluation research findings – a use beyond formative and summative evaluation for specific programmes"(1996: 134).

Patton argues for purposes of evaluation to be expanded beyond the summative and formative dichotomy in order to accommodate all purposes that come into being as the evaluation field grows.

3.5.2.4 Developmental purpose

The developmental purpose of evaluation was conceived by Patton as an evaluative response to what he calls developmental programming. This kind of programming is different from common programming because it does not take as its purpose any predetermined goals to be accomplished; rather, the process is itself an outcome. Patton writes about developmental programmes in these terms:

> "They eschew clear, specific, and measurable goals up-front because clarity, specificity, and measurability are limiting. They've identified an issue or problem and want to explore some potential solutions or intentions, but they realize that where they end up will be different for different participants – and that participants themselves should play a major role in goal-setting. The process, then, involves engaging participants in setting and achieving their own goals. Program designers observe where they end up and make judgements about the implications of what has happened for future programming and re-engineering. They never expect to arrive at a steady state of programming because they're constantly tinkering as participants, conditions, learnings, and context change. … Rather, they aspire to continuous progress, ongoing adaptation and rapid responsiveness. No sooner do they articulate and clarify some aspect of the process than the very awareness becomes an intervention and acts to change what they do" (Patton, 1993; 312-3).

Patton argues that the continuous nature of programming calls for the kind of evaluation that supports activities of such programmes, and since they are developmental, he calls this "developmental evaluation". This view of evaluation does not perceive change as progress because the purpose is not to make judgements. Change is viewed as an inevitable aftermath of evolving dynamics within and outside the programme both physical and personal, and therefore, programming has to adapt to it. Evaluation, in turn, adjusts to programming. Patton defines the "developmental evaluation process" in similar terms to other evaluations, that is, it involves asking evaluative questions and applying evaluative logic to support programmes or what is being evaluated with the evaluator playing the usual role of collaborating with stakeholders to conceptualize, design and test instruments. But in contrast to other approaches, it is 'a long-term, on-going process of continuous improvement, adaptation and intentional change'. The long-term involvement in programmes by evaluators, results in them being co-opted into programmes, hence Patton's concession that he becomes part of decision making in such programmes. The evaluator then plays a double role of an evaluator and a decision maker. Such a status is anathema in evaluation practice because of the perceived risks associated with loss of objectivity, credibility as well as the ethical implications it poses. However, developmental evaluation does not call for any form of judgement in decision making. Therefore, some of the strict requirements of evaluation do not apply to it. Ultimately, Patton summarises his views to give 'a formal definition' of developmental evaluation as follows:

> "*Developmental evaluation*: Evaluation processes and activities that support program, project, product, personnel and/or organizational development (usually the latter). The evaluator is part of the team whose members collaborate to conceptualize, design, and test new approaches in a long-term, ongoing process of continuous improvement, adaptation and intentional change. The evaluator's primary function in the team is to elucidate team discussions with evaluative data and logic, and to facilitate data-based decision-making in the developmental process" (1993:317).

Having expounded the purposes of evaluation, a question arises: Whose purposes does evaluation serve and at what levels? The following section on beneficiaries of evaluation attends to that question.

3.5.3 Beneficiaries of evaluation and spheres of evaluation utilisation

An evaluation enterprise has many stakeholders from diverse backgrounds and the attendant variations of interests in evaluations. Patton mentions *programme staff* as the most likely group to benefit from evaluations. This is so because of their daily interactions with programmes and the fact that they are at the coal face of implementation of whatever recommendations that are made by the evaluating team. Another constituency is *funders* who benefit from the findings: these permit them to make informed decisions concerning continuation of support for programmes. *Community people* can take advantage of the findings to advocate for improvements in their living conditions or gain something as a consequence of the recommendations of evaluations. *Administrators* are said to use evaluators as consultants. Conceptual use of evaluation is more pervasive in that it affects the thinking and conceptualizing abilities of *people far away from daily operations programmes*. The last constituency is an implicit acknowledgement by the author that the compass of evaluations goes beyond intended users. Of course no evaluation can serve all stakeholders equally well. Therefore, the significance of evaluation to various stakeholders varies in accordance with intentions of evaluation conduct. It is propounded that *evaluators* themselves may be beneficiaries of evaluations.

> "An evaluation may serve the needs and interests of the evaluator more than those of any other constituency. Whether such use is acceptable, justifiable or important depends on where one is sitting and what values are brought to bear on the question of utilization. Nor am I talking here simply about lining the pockets of evaluators or providing academics with publishable papers.…During the first year of operations, the evaluation unit had conducted several evaluations all of which the state administrator judged to be useful "because they helped the evaluators learn how to conduct evaluations at the state level."" (Patton, 1985: 11).

It is important to note that these evaluations were deliberately intended to generate the ability for future performance of evaluators in the new internal evaluation unit, therefore, the processes of evaluation aided evaluators. This is not to discount any other use that might have accrued to other stakeholders, but to point out that some studies essentially helped evaluators more than other constituencies.

Discussion on the spheres of evaluation is related to the level of government the needs of whom an evaluation is supposed to satisfy among different levels – the *local*, the

provincial or the *national* level. It is almost impossible for a single evaluation to cater for all of these spheres because of the often conflicting information needs and purposes for conducting evaluations. These variations have implications for designs and potential use of evaluations. "The state system imposes data collection requirements on local units that they perceive to be useless while data collected entirely by local initiative seldom meets the needs of either state or federal governments. Local units tend to prefer highly idiosyncratic and situationally specific data. Larger units tend to prefer standardized data which makes aggregation and comparisons easier" (1985: 14). The challenge is that service delivery is expected at the local level but the national and the provincial levels possess the resources to support evaluations. Evaluations commissioned by higher levels may be self-defeating when financial muscle is elevated above real issues, that is, an evaluation of an under-performing entity at a local level will not help it if it ignores its needs and context at the behest of the distant ones. This can only lead to its extinction.

We have had a taste of this in South Africa when the national government imposed a new curriculum, Outcomes Based Education (OBE) shortly after 1994. Not to be outclassed, the new minister of education introduced another one shortly after taking office in 1999. The two are deemed to have failed because of the country's poor performance on literacy and numeracy in international competitions. Now we are on the threshold of a new programme called Breakthrough to Literacy which is home language literacy learning, but many education practitioners do not know what it entails. This does not necessarily mean that it will fail. I only question the wisdom of excluding educators because they are intended users.

3.6 Misutilisation

At the end of the 1986 edition of "Utilization-focused evaluation", Patton raises some questions about the problems of misutilisation, especially in the light of perennial concerns from evaluation professionals that utilization-focused evaluation is potentially co-optive to evaluators, that is, it can turn them into advocates of programmes they are evaluating. Co-option has negative ramifications for process and findings outcomes in the sense that it casts suspicions on the neutrality of the study, thereby questioning its credibility. Patton therefore posited his preliminary observations as a way of stimulating additional discussion on the subject. These observations were once more put forward at the informal Malibu meeting of about ten evaluators and they laid the

basis of further engagement by evaluators on this issue of misuse. Prominent among them are:

> "Misutilization is not at the opposite end of a continuum from utilization. There really are two dimensions here. One dimension is a continuum from nonutilization to utilization. A second continuum is nonmisutilization to misutilization. Studying misuse is quite different from studying use. Misuse can be either intentional or unintentional. Unintentional misuse can be corrected through the processes aimed at increasing appropriate and proper use. Intentional misuse is an entirely different matter to which, it seems to me, we have paid very little attention except to say it shouldn't happen" (Patton, 1986:336).

The challenge of exploring misutilisation further has been taken up by some evaluators, among them Alkin and Coyle. In their 1988 study, "Thoughts on Evaluation Utilization, Misutilization and Non-utilization" they further developed this discussion along Patton's observations, but also discovered some twists which tend to blend the two continua. With regards to the first dimension (a continuum from utilisation to nonutilisation), *utilisation* is when a well executed evaluation is used as it should be. *Nonutilisation* is a bit complex as it incorporates an element of intentionality. It has three tiers. The first refers to evaluation findings that are unintentionally left to gather dust on the shelves - this constitutes *nonuse* of evaluation. The second layer is when a decision maker intentionally resolves to thwart the results from reaching the public; this is *misutilisation*. Depending on the nobility of the purpose, censorship of evaluation results can be warranted. This apparently blatant misutilisation can be justified to become *appropriate non-use* (*justified non-use*). Alkin and Coyle urge us to "consider intentional non-use of poorly conducted evaluations as a responsible and appropriate action. Thus misutilization in one situation may be conceived of as appropriate non-use in another" (Alkin and Coyle, 1988: 333). It is worth noting that misutilisation in this instance has actually encroached onto the use-nonuse dimension. But when the same action is taken for malicious or self-serving purposes or certain alterations are effected, *abuse* of evaluation is committed.

Another question relates to ascribing blame for misuse and whether users or evaluators should be held culpable for it. The contention is that both users and evaluators are liable. For instance, in a case of misinterpretation of results, it is implied that evaluators should have the foresight to consider users' information needs to avoid unintentional

misuse by the latter. A poor report is also the responsibility of the evaluator. Then there are issues surrounding the employment of inappropriate methodologies, flawed data collection techniques, failure to qualify results and so forth. Evaluators are supposed to have the expertise to recognize these mishaps and therefore, carry the blame for them, but since this is a case of evaluators failing to perform their duties meticulously, this is categorised under *misevaluation*. When uninformed users use information produced under such dubious circumstances, an evaluator has misevaluated, and s/he bears the brunt. But when an informed user uses it, it becomes *misuse* by her/him. The authors note:

> "Generally, when we allude to misuse, the focus is on users rather than evaluators because users are more attuned to the political ramifications of certain types of evaluations (and results), and have a greater stake in how such results are perceived. Thus, users, we believe, are more likely than evaluators to intentionally manipulate some part of an evaluation (e.g., the conclusions)" (Alkin and Coyle, 1988: 334).

It therefore seems that users are usually responsible for misuse while evaluators are liable for misevaluations. In addition to nonuse due to misevaluation, findings may be discounted because they are at variance with the preferences of certain politically powerful people and they use their power to disregard them - this is labelled *political nonuse*. *Aggressive nonuse* goes a step further by attacking the evaluator and undermining the findings that clash with established views. Patton has broken down misuse into three categories in the latest edition of "Utilization-focused evaluation". *Mischievous misuse* is when decisions or opinions are unscrupulously influenced through intentional suppression, misrepresentation or unbalanced reporting. At times information users are not competent enough to interpret findings, and end up using them inappropriately, this is called *inadvertent* or *mistaken use*. This kind of misuse can be corrected through stakeholder education. Lastly, there is *overuse*, whereby there is a discrepancy between the findings and the importance attached to them. In this instance too much importance is given to findings.

Patton draws a distinction among types that are commonly lumped together as one or used interchangeably – political, symbolic, legitimative and persuasive. He categorised them under *political uses* as the umbrella term because power relations is the common denominator among them. *Symbolic use* is when an evaluation is supported only to

make it appear as though decisions are based on empirical evidence. An evaluation may be commissioned to support a decision that has already been taken, this is called *legitimative use*. An additional use is what Weiss, Murphy-Graham and Birkeland called *imposed use* in "An Alternate Route to Policy Influence: How Evaluations Affect D.A.R.E".

> "In these stories you will see instrumental, political and conceptual use, and also a new type of use, previously unaccounted for – probably because it is making a debut on the evaluation scene. …it is a type of use that came about because of pressure from outside. In this article, we adopt the term *imposed use*, with an occasional nod to carrot-and-stick use. What distinguishes this type of use is that program stakeholders are obliged to pay attention to evaluation results. In this case, they would lose their funding if they did not adopt a program that had been proved effective through scientific inquiry" (Weiss et al., 2005:16).

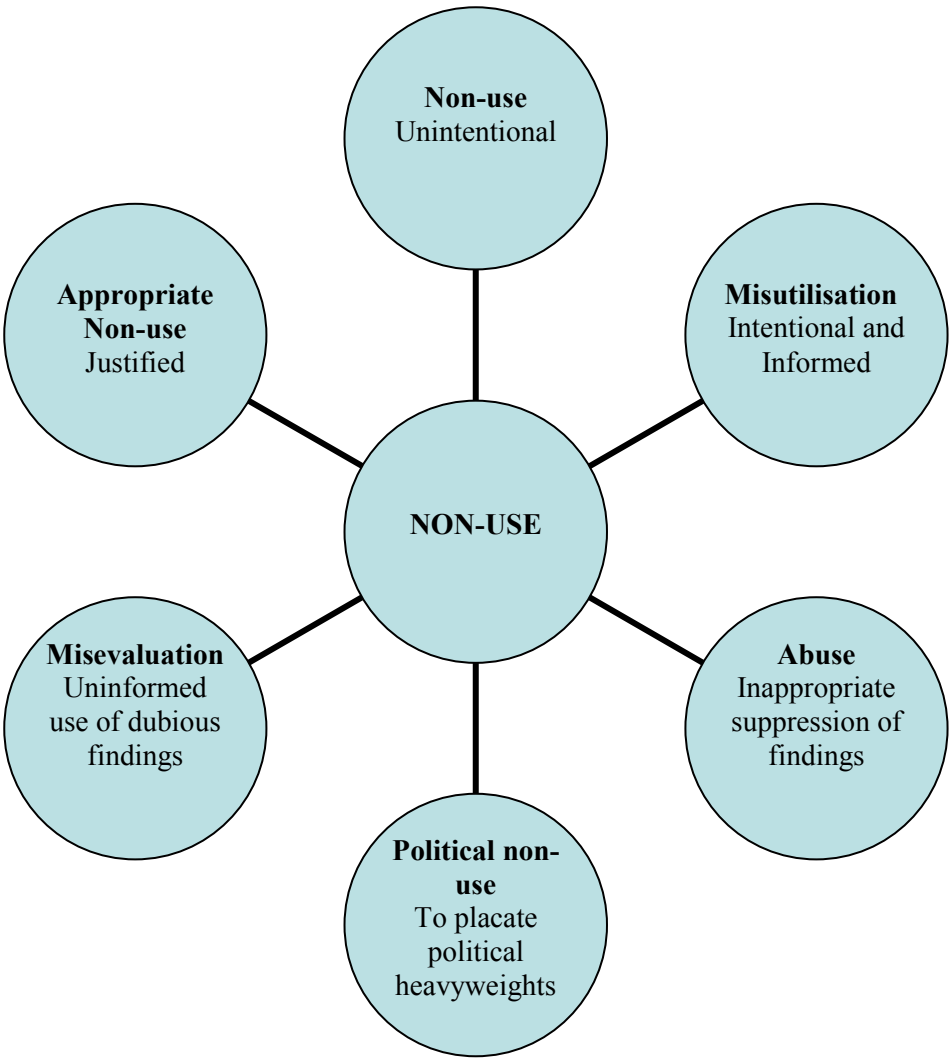
Persuasive use is the selective use of evaluation findings to strengthen one's position in political debates and funding decisions. Persuasive use is not necessarily negative because in certain circumstances it is an essential component of duty. Henry in "Why not Use?" (2000) laments the negative connotations associated with persuasion as a type of use. He declares that it has been tarnished by its links to legitimation and symbolic use. These imply that evaluation data are used publicly to support decisions that have already been taken on different grounds or that evaluation simply represents the opinion of the decision maker. He prefers Majone's definition of persuasion, which he claims, is relevant for the pursuit of social betterment.

> "To decide, even to decide correctly, is never enough in politics. Decisions must be legitimated, accepted, and carried out. After the moment of choice comes the moment of justification, explanation, and persuasion. Also, policymakers often act in accordance with pressures from external events or the force of personal convictions. In such cases arguments are needed *after* the decision is made to provide a conceptual basis for it, to increase assent, to discover new implications, and to anticipate or answer criticism" (Majone [1989] as cited by Henry, 2000: 92).

This explanation clarifies and justifies the chronology of events – decision followed by evaluation – in persuasion, thus, removing the negativity surrounding the purpose of evaluation. Henry supports Majone in her pronouncement that evaluation is not well placed as a formal demonstration of what should be done; rather, it is a 'tool for

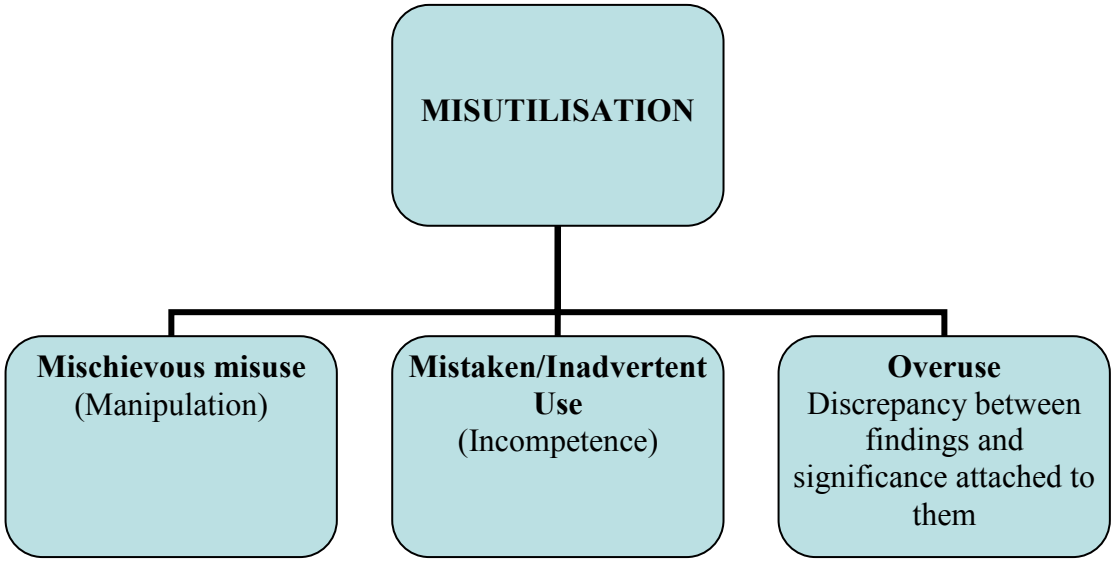
argumentation'. Evaluation, she contends, 'does not conclusively demonstrate the choice that should be taken, but it provides justification and support for the existence of a problem or a course of action'. He mentions that the *paradox of pursuing persuasive use* is that "when the evaluation is planned, designed, and executed to facilitate persuasion as the particular form of use, the evaluator must begin with the objective of producing persuasive evidence. The endeavour becomes not evaluation but "pseudo-evaluation", or an exercise of political advocacy, a well-recognized but roundly discredited approach to evaluation" (Henry, 2000: 92). This is because if persuasion is the goal, the evaluation will pick and concentrate on those data that support the intended user's preferred policy positions to corroborate them because decisions have been made already. This can deviate from social betterment when information is only revealed when it justifies preconceived courses of action. Other information that does not support persuasive use may be dropped – this demonstrates the shortcoming of pursuing planned use as the goal of evaluation.

Figure 3.2 Typology of non-use



Lastly, misutilisation can be that of *commissioning an evaluation*, when an evaluation is commissioned solely for political/symbolic reasons, *misuse of evaluation processes* by using evaluations to delay action and *misuse of evaluative findings* – shelving results, selective reporting, uncritical consumption of findings *etc*.. Bradley J. Cousins rehashes Alkin and Coyle's framework in "Commentary: Minimizing Evaluation Misuse as Principled Practice" (2004). He takes the discourse further to suggest some mechanism for dealing with such mishaps by relying heavily on the provisions of the Guiding Principles for Evaluators (American Evaluation Association) for solutions. Different imaginary scenarios are created and readers are taken through the guiding principles quoting relevant sections and clauses from them for particular circumstances evaluators might find themselves in. It is more like a law academic guiding students through a maze of statutes for pertinent pieces of legislation for various case scenarios. It does not help much except to alert novice evaluators to the existence of guiding principles for evaluation and how they can help.

Figure 3.3 Typology of Misutilisation



3.7 Conclusion

The definitions of evaluation research are varied owing to the broad spectrum of fields in which evaluation is conducted and the variations in emphases follow the foci of those different professions. The differences are more pronounced because of the secondary nature of evaluation to the purposes of the professions in which it was carried out, before it became a fully fledged profession. For example, in education, evaluation plays second fiddle to education; the same pertains in psychology and to a host of other professions. The definition evolved from being judgmental about programmes in the first edition of "Utilization-focused evaluation" (1978) to a stage where evaluation research was separated from programme evaluation in an almost dichotomous relationship in the second edition (1986). The significance of the definition lies in its alignment with ideas in Patton's writings so that it becomes easy to understand Patton's discourse due to the knowledge of the context in which he applies evaluation research, which comprises the bulk of his works. With regards to use, Patton relegates the responsibility for its definition to intended users. Nevertheless, readers get a broad idea of what it involves from its typology and the purposes of evaluation. The types of use are a mixture of influence on and originality from him as indicated by inclusion of uses such as instrumental, conceptual and political which predate his writings on them. His major contribution to the typology was the discovery of process use in the 1990s which has since become the subject of many scholarly papers. Process use, just like findings of evaluations, has developed to have its typology. That is where Carol Weiss' contention comes in - there is an anomaly when the use of findings is not regarded as findings use, but the use of processes is considered as process use. She propounds that process use is not the type of use but a source of use. Nonetheless, the contention does not detract attention given to process use.

The other major observation on use by Patton was the separation of use from misuse by suggesting that the two concepts are not the opposite ends of a single continuum. That led to the development of a typology of misutilisation by evaluation researchers and practitioners like Alkin and Coyle, Burry and Cousins among others. Patton also added other evaluation purposes to the popular formative and summative evaluations in the form of developmental evaluation and knowledge generation and so forth.

CHAPTER FOUR

PATTON'S INVOLVEMENT IN THE PARADIGMS DEBATE

4.1 Introduction

The rise of logical positivism in the middle of the twentieth century coincided with the corresponding ascendance of the quantitative paradigm which employed experimental and quasi-experimental methods. In social sciences research and practice experimental domination manifested itself in the works of proponents like Campbell and Stanley (1963), Cook and Campbell (1979) which were renowned at the time and propagated experimental designs as *the* model of research, while rating methods related to the qualitative paradigm very poorly. These theorists and a host of others firmly believed in the unquestionable importance of experiments in determining cause and effect and establishing general laws for human behaviour. This domination of the quantitative paradigm over the qualitative one culminated in the so-called the paradigms debate between these two approaches to evaluation studies in the 1960s and 1970s. Patton's involvement in the debate in the 1970s on the side of the naturalistic paradigm was also a response to the dominance of experimental methods in evaluation.

4.2 Early reaction to the dominance of the quantitative paradigm

Patton was exposed to the qualitative paradigm through the North Dakota Study Group on Evaluation. The purpose of the study group was to discuss the shortcomings with evaluation as it was applied at the time. Vito Perrone, the facilitator of the study group and the Dean of the Centre for Teaching and Learning, University of North Dakota propounded that educators across the length and breadth of the United States of America had met at the university to "discuss issues of common concerns about the narrow accountability ethos that had begun to dominate schools and to share what many believed to be more sensible means of both documenting and assessing children's learning" (Patton, 1975: 5). Considering the fact that evaluation was a common practice, albeit predominantly quantitative, in United States schools and that it was supposed to curb undesirable practices in those schools observed, the "unaccountability ethos" referred to was a clear critique of the experimental paradigm. That might have planted the seeds of doubt in the dominant paradigm in Patton's mind, and judging by his works on qualitative paradigm, the seeds took root and germinated into a strong inclination towards the latter paradigm. His paper on "Alternative Evaluation Research Paradigm**"** (1975), was an early reaction to the domination of the quantitative approach

over the qualitative one in educational and social evaluation research. It was also a response to an increasing public clamour for effective programmes in those disciplines, and programme evaluation is an important instrument in that regard. That necessitated a change in the manner in which things were done. Patton declares that "what makes consideration of an alternative research paradigm so pressing is the fact that *these prominent exemplars of evaluation are largely based on a single, largely unquestioned, scientific paradigm*"(Patton, 1975: 7). He cites the 1974 extensive study conducted by Bernstein and Freeman on evaluation of evaluation research an example of scientific its dominance. This study was funded in the 1970 fiscal year by the government and based on 236 research projects using experimental designs.

The purpose of the study was to assess the quality of evaluation research projects anchored in six "indicators", namely: the nature of research design, representativeness of the sample, sampling, type of data analysis, nature of data analysis and quality of measurement. It is worth noting that the kind of quality variables selected, with references to representativeness and measurement, are a thinly veiled bias towards the quantitative paradigm; therefore, it is not unanticipated that qualitative variables were inadequate as quality indicators, they were intended to be. In defence of this unashamedly prejudiced study, the authors cite the works of others who share the same bed with them which "strongly suggest that the best evaluations in terms of research quality are those which are highly quantitative" (idem: 10). Having been exposed to the qualitative paradigm, Patton did not react well to this kind of exclusive logic where evaluation research was equated with scientific approach and touted as the only route to social progress. He consequently set out on a mission to propagate an **alternative** paradigm that draws on qualitative methodology and predicated on the notion of the doctrine of *Verstehen*, phenomenology and ethnomethodology. According to him qualitative methods are 'derived most directly from the ethnographic and field study traditions in anthropology'.

The ***Verstehen*** doctrine acknowledges the differences between human behaviour and that of other objects and as such, human behaviour cannot be researched and understood in the same manner as objects are studied and appreciated. According to Strike (1972) as quoted by Patton, the *Verstehen* doctrine states:

> "Men have purposes and emotions, they make plans, construct culture, and hold certain values, and their behavior is influenced

> by such values, plans, and purposes. In short, a human being lives in a world which has "meaning" to him, and, because his behavior has meaning, human actions are intelligible in ways that the behavior of nonhuman objects is not"(Patton, 1975: 13).

It is apparent that this approach reacts against the relating together of human behaviour and the sciences that study it with physical objects and their sciences (positivism). Human behaviour cannot be subjected to certain preconceived laws whereby those actions that do not fit into those laws are discarded as the influences of extraneous variables. Verstehen is an understanding that focuses on meaning of human behaviour, taking into consideration contextual factors.

Ethnography also had an influence in moulding Patton's approach to alternative paradigm. Ethnography is the method of study commonly used by anthropologists in their human behavioural studies. It is an in-depth approach that studies 'entire groups that share a culture or sub-cultures within bigger sub-cultural institutions with the purpose of identifying cultural patterns such as norms, traditions and so forth. The method used is usually site-based fieldwork where the ethnographer is immersed in a cultural group they would like to investigate and observe and record their activities in their natural setting. Martyn Hammersley traces the philosophical foundation of ethnography with its emphasis on 'naturalism' (capturing social phenomena in their own terms) to realism. He writes that:

> "a doctrine of realism, by which I mean that there is reality independent of the researcher whose nature can be known, and that the aim of research is to produce accounts that correspond to that reality. There can be little doubt about the widespread acceptance of this view. It is a philosophical doctrine on which much ethnography is founded" (2002: 66).

The researcher has to be as unobtrusive as possible and must not have a vested interest in the outcome of the study s/he is conducting. The greatest strength of ethnography is its faithful representation of social phenomena. The researcher is also able to verify interview responses against observations of the group's activities in a natural setting and is better placed to interpret their behaviour more accurately than in an interview having observed them for a considerable time - months or years, and therefore, has a better understanding of their world. However, ethnography is not without limitations. There is always a lurking danger that the researcher may go "native", that is, the researcher may unconsciously change from being an outsider to an insider, get

emotionally involved and then lose the ability to assess the situation objectively or accurately. It is also time consuming.

Phenomenological studies are concerned with people's perceptions, perspectives and their understanding of a particular situation. It is more dependent on interviewee explanations of what they feel about their experiences than the researchers' judgements of them. "In some cases, the researcher has had personal experience related to the phenomenon in question and wants to gain a better understanding of the experiences of others. By looking at multiple perspectives on the same situation, the researcher can then make some generalizations of what something is like from an insider's perspective"(Leedy and Ormrod, 2005:139). The most used method of research in phenomenology is interviews, and like other qualitative interviews, phenomenological interviews are open, long and unstructured. The researcher (interviewer) and the respondent work together to find the answer to a research question. This is because of the nature of the research problem – people's perceptions of situations. Respondents do most of the talking while the researcher listens and records. As is the practice with qualitative studies, the researcher must take care not to be involved and lose objectivity.

First propounded by Garfinkel, **ethnomethodology** is part of an ongoing debate about the compatibility of natural science methods with the social sciences and it takes the interpretivist tradition a step further. It is concerned with exploring ways in which people interact with the world around them and attempt to make sense out of it while maintaining order through their interactions. "A central question that the ethnomethodologist asks is: how is it that everyday behaviour appears on the surface level to be stable and ordered even in situations where there is potential for misunderstanding and ambiguity?"(Babbie and Mouton, 2001: 30). It is founded on the conviction that people are not passive reactionary objects which react to stimuli of the environment, but rather, they are proactively involved in shaping it, bringing about orderliness to it through certain 'tacit' rules on what is acceptable and what is not acceptable in their society. The aim is to understand those social orders that people employ to make sense of their reality by analysing their own accounts and describing their 'day-to-day' experiences. Like qualitative approaches to research, ethnomethodology does not pass judgements on human behaviour nor strive for causal explanations. The main emphasis is on *how* they make sense of whatever reality they find themselves in on a continuous basis.

Patton's ultimate reaction to the dominant paradigm was his participation in the paradigms debate where he put a case for the alternate paradigm, showing its strengths and contrasting it with the dominant scientific approach in the form of dichotomous dimensions. However, he concedes that the paradigms are not as opposed as the dichotomies indicate; they are only meant to facilitate the paradigms' dialectics.

Patton argues that whereas the scientific paradigm is preoccupied with reliability the alternative paradigm emphasises validity. Reliability refers to the consistency of scientific findings, whether or not they can be replicated to other environments with similar characteristics. There is also an implicit assumption that there is a correlation between reliability and validity, that is, reliability automatically leads to validity in a linear kind of relationship. However, Patton cautions through the words of Irwin Deutscher that "we concentrate on consistency without much concern for what it is that we are consistent about or whether we are consistently right or wrong. As a consequence we may have been learning a great deal about how to pursue an incorrect course with a maximum precision" (Deutscher [1970] as cited by Patton, 1975:24). While it may be true that the absence of reliability will almost certainly result in a corresponding lack of validity, its presence does not necessarily guarantee validity. Therefore, attention must be given to using methods that reduce distortions of findings. Standardised tests may be reliable and consistent, but not inevitably valid measures of learning in all situations. A quest for reliability should not blind researchers to contextual factors. Patton writes about operational definitions and measures of major educational and social scientific concepts thus:

> "One sees that their transparency and bias are frequently astounding though their reliability is extremely high. In addition, we seem to have lost sight of the fact that responses *mean* different things in different settings and different contexts. (The only way to discern such variations in shades of meaning is to directly interact with and observe respondents in various relevant settings.) Thus, instruments prepared for evaluation in one setting are adopted for evaluation in other settings with a facility that shows arrogant insensitivity to the issue of cross-setting validity" (Patton, 1975: 25).

Validity in the alternative paradigm is about 'meaning and meaningfulness of data collected and instrumentation employed', that is, verification that instruments really measure what they are supposed to measure and that the meaning of data is not distorted. These can be assured by developing new instruments for each evaluation and

ensuring that they are valid for the setting in which they are used. By getting close to data and empathising with participants, alternative qualitative evaluators are able to understand their perspectives and the meaning of data to actors themselves, thus giving representation of findings in their perspectives.

4.3 Opposition of objectivity to subjectivity

Within the positivist tradition, objectivity is greatly revered in experimental methods as the hallmark of reliability and impartiality in direct opposition to subjectivity which is regarded as personal intuition and value laden, and therefore not worthy of the name science. Qualitative research, on the other hand, is considered by adherents of the scientific tradition as subjective or without consistent measurements, and therefore biased. The difficulty with the notion of objectivity is that in some instances evaluation may be intertwined with politics to the extent that it becomes impossible to eliminate subjectivity. Nonetheless, the scientific approach employs its procedures and techniques to 'control' such subjectivity. Patton asserts that in such cases 'control' is a euphemism for evading relevant political issues by limiting and prejudicing questions that can be asked to arrive at certain outcomes that are amenable to statistical manipulations. This leads Patton to conclude that the dominant paradigm's insistence on itself as the sole preserve of objectivity is an ideological statement meant to 'legitimize, preserve and protect the dominance of a single evaluation methodology'. While conceding that qualitative evaluation research is less restrictive on respondents and evaluators in terms of engagement, he rejects the insinuation that it is not objective and the suggestion that the traditional paradigm is unfailingly objective. He enlists the words of Michael Scriven to illuminate this point. Referring to this misrepresentation of facts, Scriven (as cited in Patton (1975) writes:

> "Errors like this are too simple to be explicit. They are inferred confusions in the ideological foundations of research, its interpretations, its applications. …it is increasingly clear that the influence of ideology on methodology and of the latter on the training and behavior of researchers and on the identification and disbursement of support is staggeringly powerful. Ideology is to evaluation what Marx suggested the economic factor was to politics and what Freud took sex to be to psychology" (Patton, 1975:28).

Patton agrees with Scriven that science has led to a formalistic split between the mental and the logical, seen as the subjective and the objective. This split has the effect of preventing researchers from realizing that understanding, properly conceived, is in fact

an objective state of mind or brain and can be tested quite objectively (idem:28). The point here is that subjectivity need not be opposed to objectivity and when it is interpreted in this manner, subjectivity can contribute positively to research by enabling the research to get close to data, "thereby developing analytical, conceptual, and categorical components from data itself - rather than from a preconceived, rigidly structured, and highly quantified techniques that pigeonhole the empirical social world into the operational definitions that that the researcher has constructed" (Filstead [1970] as cited by Patton 1975:29). Getting close to data also serves to remind scientists that the significance of science lies in its pragmatism, that is, its ability to shed light on the solutions to critical social problems. That is done without mediation of any premeditated theoretical categories and abstractions of numbers in the alternative paradigm as opposed to the scientific paradigm where there is prior interference and manipulation in the form of controlled stimuli.

The matter of the evaluator's proximity to data is a big bone of contention between the dominant scientific and the alternative naturalistic paradigms. The former propagates detachment from data as a way of maintaining neutrality and objectivity. These supposedly manifest professionalism in as far as they avoid personal involvement - they let skills do the work. A question arises, are skills without personal understanding able to serve the needs of clients? Patton argues that the persistence of this call for distance from data can be interpreted as an endeavour by scientific evaluators to account for their "alienation" from the social world in search of 'status and professional prestige' and elitism associated with scientific revolution by creating laws for prediction of human behaviour. Horowitz as cited by Patton states:

> "This alienation is built on the notion that society and culture can be viewed like any other 'natural' phenomena, as having laws that operate quite apart from the intentions, motivations, and plans of human beings. The methodology follows this assumption by emphasizing prediction and universal laws rather than understanding and human meaning" (Patton, 1975:32).

According to Patton, the alternative paradigm, on the other hand, questions the necessity of distance from data because it seeks to understand human behaviour, not quantify it. Its proponents contend that understanding is predicated on empathy and sympathetic introspection that derive from the observer's personal involvement. The qualitative researcher is willing to get close to data and to capture social actors 'in their own terms' to avert preconceived ideas about results of studies. While not discarding

the significance of statistical analyses, a qualitative researcher argues that data must be interpreted in an appropriate manner and be given a similarly relevant human meaning in the terms of human actors studied. That requires personal involvement. Lofland, argues that "in order to capture the participants 'in their own terms', one must learn *their* analytic ordering of the world, their categories of rendering explicable and coherent the flux of raw reality" (1971: 7). The alternative paradigm also finds a contradiction in the traditional approach's claim to knowing human beings between their daily personal dealings and judgements of other human beings on the one hand, and their insistence on the superiority of statistical depictions, on the other. In everyday life they reason like everyone else, assuming that they do not know or understand well people they do not associate much with them. They accede to the fact that in order to know and understand people, they have to be acquainted and interact with them 'face-to-face' in various environments reasonably often. But once they wear their scientific evaluation caps, they somersault and discard the notion of 'face-to-face' interaction as a prerequisite for knowledge and understanding.

Patton also raises the critical issue of a match between research methodology and research problem. Instead of a one fits all methodology, researchers should strive for appropriate methods for research problems – a personalized intervention for a personalised environment. Patton concludes by reminding experimental pundits that many studies of great eminence in the 'real' scientific world (physical sciences) they strive to emulate have come from personal experiences. These include "Piaget's closeness to his own children, Freud's proximity to and empathy with his patients, Darwin's closeness to nature, and even Newton's intimate encounter with an apple" (1975: 34). The logic of the dominant paradigm spells the danger of throwing these laudable contributions to human and scientific progress out of the window of scientific research while elevating them to pseudo-science.

4.4 Debate about designs

Experimental designs commonly deal with a few sets of variables that include the treatment/independent variable and the dependent variables. Causes and effects have to be operationally defined. These do not always apply smoothly in educational and social sciences as Patton points out that:

> "One of the major problems in experimental educational research is clear specification of what the treatment actually is, which

> infers controlling all other possible causal variables and the corresponding problem of multiple interference and interaction effects. It is the constraints posed by controlling the specific treatment under study that necessitates simplifying and breaking down the totality of reality into small component parts" (1975: 35).

The fact that the scientific method is predicated on simplification of the studied reality should be cause for alarm about possible distortions of that very reality. Moreover, the parochialism of the focus of experiments, artificial controls and isolated treatments should be of further concern about the accuracy of results. One may justifiably argue that the findings of experimental methods are predetermined through these interferences and manipulations. There is also an issue of relevance of laboratory experiments to wider social settings being evaluated, where actions happen spontaneously. There is an apparent disjuncture in the two 'realities', one being artificial and the other real. Yet the scientific approach is adamant in proclaiming itself the superior and reliable vehicle to empirical knowledge. The alternative paradigm, on the other hand, contends that for evaluations to be meaningful and relevant to contexts in which events occur, they have to include 'a holistic methodological approach built on the functioning, day-to-day world of program participants' (idem: 36). The most irksome problem to Patton was the widespread employment of experimental designs at that time to 'all encompassing' innovations like open education that involves changes in curricula, materials, teaching methods which also changed the social relationships among students and between students and educators. He vehemently argues that such an interactive approach and the interdependence it engenders make it impossible to isolate and control its component parts. Therefore, an alternative holistic methodological approach is necessary to capture and describe in meticulous detail the meaning of classroom interactions from students' and teachers' perspectives. A holistic approach transcends artificial conflicts of modern schools grounded in academic divisions which John Dewey observed were 'alien to the nature of the child'. Patton reproduces Dewey who wrote:

> "Again, the child's life is an integral, a total one. He passes quickly and readily from one topic to another, but is not conscious of the transition or break. The things that occupy him are held together by the unity of the personal and social interest in which his life carries along… His life is fluid and fluent; its contents dissolve and re-form with amazing rapidity" (Dewey [1966] as cited in Patton, 1975: 36).

The dominant experimental paradigm continues to ignore this totality of personal experiences by assuming that certain isolated variables influence human behaviour. It is therefore not surprising that Patton notes that teachers aired their disenchantment with the bulk of experimental evaluation findings lack of relevance to innovative programmes.

4.5 Focus of evaluation

The experimental model is premised on the assumption that there is one isolated and measurable treatment that yields results, hence the preoccupation with outcomes. Experiments are conducted prior to (pre-tests) and after (post-tests) the treatment and the results of the two tests are compared and contrasted. The underlying assumption for this practice is that the administered treatment remains constant between the tests and that it accounts for change in the treatment group. Patton admits that there are some educational programmes which fit this narrow description, but the fluidity of most educational innovations does not permit this parochial approach. The final implementation is usually different from the original innovation because they undergo metamorphosis as educational practitioners learn what works and what does not work during implementation. These changes upset scientific evaluators who seek scientific, constant treatments to static programmes in order to arrive at predetermined outcomes. Reactions towards these evolving innovations are usually negative and the consequent evaluations do not advance the noble purposes of evaluation as demonstrated in the assertion hereunder:

> "Because of commitment to a single evaluation paradigm evaluators are frequently prepared to actually do everything in their power to stop program adaptation and improvement so as not to interfere with their research design" (Patton, 1975:39).

It is not far-fetched to argue that the presence of an evaluator in such circumstances defeats the very purpose of evaluation – social betterment through effective programmes. Perhaps the rationale of the scientific evaluator is that any kind of improvement that is initiated by programme staff during evaluation is inferior, and therefore must be sacrificed at the altar of superior scientific experimentation, which often yields unintelligible findings to intended users. According to Patton, the kind of arrogance displayed in such cases is actually discouraging and has repercussions for the experiment itself. If the drive for constancy of treatment results in low staff morale and

concomitant less participation, then treatment will be under different conditions from the environment which necessitated evaluation in the first place. Treatment will change in line with changing conditions and therefore contaminated. In extreme cases, the researcher effects may have unanticipated consequences of actually worsening the situation than that of the control group because of low staff morale in the treatment group. Programmes are not initiated and tailored to suit the needs of evaluators. Therefore, it is not the task of programme administrators and other stakeholders to appease evaluators. On the contrary, administrators strive to run programmes in ways that best achieves their goals even if it means effecting changes during evaluation. And it is the evaluators' task to come up with research designs that fit programmes, not the other way round. The alternative qualitative evaluation approach is appropriate for innovative programmes because of its emphasis on processes. Patton writes that:

> "Process evaluation is not tied to a single treatment and predetermined goals or outcomes. Process evaluation focuses on the actual operations of a program over a period of time. The evaluator sets out to understand and document the day-to-day reality of the setting or settings under study. Like the anthropologist, the process evaluator makes no attempt to manipulate, control, or eliminate situational variables or program developments, but takes as given the changing reality" (1975: 40).

In qualitative research the focus is not only on the outcomes, but also on the processes and therefore most likely to detect changes in treatments and reactions necessitated by changing reality. The changes themselves can serve to improve programmes – process use.

4.6 Contextual inimitability and generalisation

The experimental and quantitative emphasises the establishment of laws and theories. These are done by uncovering patterns of behaviour and their causes so that it becomes possible to predict behaviour from action. This allows scientists to postulate laws for human actions. Patton is concerned that it is now taken for granted that the purpose of science is to produce such generalizations because they are not questioned. He claims that 'philosophers suggest that the greatest contribution of Western culture and civilization is the value it places on the individual' (1975: 41), yet proponents of the scientific approach ignore this basic tenet of liberal philosophy and its contribution to individuality of human beings in pursuit of generalizations. For Patton, individual

emphasis has important implications for humanistic evaluation research, especially because of the abundance of educational innovations and social programmes that focus on individual clients.

In educational evaluation research, the pursuit for generalizations distracts teachers from paying attention to individual learners as needs arise in classroom situations because education is both idiographic and nomothetic. This individuality is in fact recognized by educational researchers in their discussions of "disordinal interactions" (treatments interacting with personological variables in educational experiments) that account for learners' variations in responses to innovations. According to Patton, experimental researchers seldom uncover these disordinal interactions. The point is driven home by Bracht and Glass (1968: 451) who wrote:

> "In point of fact, the actual search for disordinal interactions is very rare – most researchers don't bother with the difficult statistical analyses necessary or don't measure relevant variables – and "the molarity" (as opposed to the molecularity) of both personological variables and the treatments incorporated into many experiments may tend to obscure disordinal interactions which might be observable when both the variables and the treatments are narrowly defined".

Bracht and Glass ultimately dismiss the search for these interactions as fruitless because scientific designs have trouble in capturing them in most cases, instead of questioning the methodology that fails to unearth them. It is an open secret to teachers that individual differences in learning among students do exist, but experimental designs consistently fail to discover them. This may be cause for loss of trust in or outright disdain for evaluation among educators. Patton calls for an alternative paradigm in evaluation whose methodology recognises the uniqueness of individuals and therefore will take into account the need for individualisation of teaching or the needs of individual clients in social programmes, as well as generalisations by acknowledging similarities among people. By getting close to data, being descriptively factual and quotive, the alternative approach can put forward individual cases in their own terms. Patton also laments the emphasis on large samples by the dominant paradigm in its search for generalisations. These, he argues, result in an increase in the number of cases in research studies that encourage distance from data and its quantification.

4.7 Evaluation for Whom and for What?

Ideally, evaluation research should serve the interests of both scientists and practitioners, but the dominant paradigm at that time (hypothetico-deductive paradigm) was mostly concerned about developing outcomes that would be acceptable to scientific practice. Several reasons accounted for that emphasis on scientific designs, namely: career considerations, prestige associated with science, professional commitments among others. Patton asserts that:

> "The nature of funding in most major evaluative research reinforces this emphasis by rewarding grandiose designs, elegant sampling, and sophisticated quantitative methodological procedures. Such evaluations – frequently national in scope – focus on outcome assessment and summative evaluation. *Such evaluations are virtually useless to practitioners in individual programs*" (1975: 44).

It is argued that for evaluations to be useful to specific practitioners, they must be localised and be descriptive of daily events, context, treatments and outcomes in ways that are understandable and meaningful to practitioners, and the alternative research approach offers that kind of option with its emphasis on qualitative methods. Qualitative evaluation is able to focus on individual programmes in order to identify shortcomings and strengths with a view to improve them. Unlike in large scale evaluations, stakeholders in individual programmes can exchange feedback that results in development, with evaluators. Moreover, the alternative approach can also contribute towards generating a body of scientific knowledge through accumulation of documentation that can help clarify issues of measurement, treatments and outcomes in social research.

Patton (1985) makes a call for transcendence of pragmatism over paradigmatic affiliation. Patton concurs with Bednarz and Guba and Lincoln that the scientific and the naturalistic paradigm contain incompatible assumptions about the nature of reality, but he differs with Bednarz when it comes to the implications of such incompatibilities. On the one hand, there are those who hold that since the two paradigms are so opposed, one cannot employ methods commonly associated with them in one evaluation study. On the other hand, Patton and others believe in the ability of flexible, responsive evaluators to move back and forth between the paradigms in a single evaluation. Guba also argued that the issue of holistic-inductive analysis in naturalistic research is not

absolute because the researcher usually oscillates between a 'discovery mode' and a 'verification mode'. Patton (1990: 46) confirms that "as the research begins the investigator is open to whatever emerges from data, a discovery or inductive approach. Then, as the inquiry reveals patterns and major dimensions of interest, the investigator will begin to focus on verifying and elucidating what appears to be emerging – a more deductive approach to data collection and analysis". He professes to have been in situations where clients affiliated to both paradigms were split on the kinds of designs they envisaged for evaluations. Predictably, adherents of 'progressive education' demanded a naturalistic approach to evaluation of open education innovation while devotees of 'traditional education' wanted rigorous scientific methods. Whilst Patton believes in the inconsumerability of paradigms, he suggests that evaluators are better able to deal with the situation by first engendering tolerance of diversity of perspectives and then focusing on information needed by both groups and negotiate designs based on that information. He writes:

> "Multiple methods and multiple measures will give each some of what they want. The naturalistic paradigm will want to ensure that test scores are interpreted within a larger context of classroom activities, observations, and outcomes. The scientific paradigm educators will likely use interview and observation data to explain and justify test scores interpretation" (Patton, 1985b: 307).

In this way, groups can agree on the employment of both quantitative and qualitative data while paradigms debate continues. Moreover, Patton brings to our attention that paradigms are not the only incompatible aspects in evaluations. There are sensitive matters that include politics, culture, religion among others on which evaluation stakeholders do not agree, but those are overridden by the centrality of evaluation. There is a convergence of views between Patton and Guba and Lincoln on the fact that 'common commitment to empirical evaluation can serve as a bridge' among evaluators of different paradigmatic persuasions. In the end pragmatism should be the determining factor.

Patton asserts that the discussion of the alternative paradigm in opposition to the traditional one is meant to increase options on the menu of methodological paradigms, not to replace one paradigm with the other, nor a statement that there are only two paradigms. But he cautions against assumptions that have misled evaluation

practitioners into believing that paradigms are mutually exclusive. Reichardt and Cook pointed out these misconceptions:

> "First, it is assumed that a method-type is irrevocably linked to a paradigm so that allegiance to a paradigm provides appropriate and sole means of choosing between method-types. … Second, the qualitative and quantitative paradigms are assumed to be rigid and fixed, and the choice between them is assumed to be the only choice available" (Reichardt and Cook, 1979:10-11).

Patton distances himself from these suppositions, instead, he indicates that the conclusion that qualitative and quantitative methods cannot be used together based on the paradigmatic foundations is a consequence of treating 'descriptions' as 'prescriptions' – paradigms are descriptive, he argues. Patton declares that from the utilization-focused evaluation perspective, no paradigm is better than the other. They are only alternatives from which evaluators can choose. Their descriptions are aimed at sensitising evaluators to the fact that 'methodological prejudices, derived from their disciplinary socialization experiences, may reduce their methodological flexibility and adaptability' (Patton, 1980: 220) In effect, discussion of paradigms makes evaluators aware of their predispositions, thus, helping them to transcend them.

4.8 The Paradigm of choices

The ferocity with which the paradigms debate raged in the 1970's and 1980's belied an assumption that there is a single best evaluation design. But most contexts are unique in variables as well as in stakeholders, and it is almost impossible to serve all stakeholder groups equally in a single evaluation. Stakeholder diversity implies that what is best for one stakeholder group may not necessarily be so for others, therefore, "being best" is particular, that is, it is commonly the perception of the most served group in a multi-stakeholder evaluation. This, Patton argues, affords evaluators with some freedom because it casts away notions of the '*right* design, the *right* analysis and the *right* findings'. They can focus their attention on the situation, and that in turn, will determine the design, analysis and findings. That is effectively shifting from adherence to paradigms to situations – matching methods to situations/contexts. Situational responsiveness requires an evaluator in collaboration with stakeholders or intended users in utilization focused evaluation to consider a variety of options, but this is sometimes difficult because of influences of paradigms. Since paradigms shape one's worldview, as alluded to earlier, it is not easy to break free from their stranglehold,

especially when there are numerous and complex choices to be made. Some researchers simply fall back to standard operating procedures as dictated to by paradigms they were socialised in, predominantly what Patton following Kuhn refers to as 'normal science'. He writes:

> "The practice of 'normal science' within any of the major disciplines is quite different from the practice of creative evaluation. Creative evaluation is a paradigm of choices. It recognizes a broad range of viable research methods and approaches. Creative evaluation involves situational responsiveness, methodological flexibility, conscious matching of evaluation approaches to the needs and interests of those with whom one is working, and sensitivity to the unique constraints and possibilities of particular circumstances" (Patton, 1980 b: 24).

Paradigm of choices is contrasted with 'normal science' in a manner similar to disparities between scientific and naturalistic approaches. On the one hand, paradigm of choices is a creative problem-solving approach to evaluation that is centred on the evaluator's ability to adapt an evaluation to changing environment. On the other hand, there is 'normal science' with its preconceived models unto which evaluation designs are supposed to fit after subjection to manipulations. To the extent that the paradigm of choices is situationally responsive, it is actually an advocacy for the use of mixed methods, but to the degree to which it is polarized with normal science, it is tilted towards qualitative paradigm. It appears as though it is the converse of the situation in mixed methods where qualitative methods are employed as accessories to quantitative ones, the latter being regarded as the 'gold standard'. More interestingly, it is inseparable from utilization-focused evaluation in that what goes for a paradigm of choices also goes for utilization-focused evaluation – methodological flexibility, situational responsiveness, matching approaches to the needs of intended users and so forth.

4.9 Quality and credibility in qualitative evaluation studies

The notions of quality and credibility are intertwined because judgements of quality have an impact on perceptions of credibility. Conversely, perceptions of credibility may affect quality, either positively or negatively, by focusing attention on or away from it. There are diverse approaches to qualitative inquiry and each approach is tailored according to contextual needs and purposes of research. Criteria for judging quality also ought to vary dependent on purposes and situational variations. Patton (2002: 226)

highlights the importance of acknowledging that 'competing philosophical underpinnings and theoretical orientations will generate different criteria for judging quality and credibility'. He identifies five sets of criteria for judging quality in qualitative evaluation studies as perceived from different philosophical perspectives that also correspond with the development of qualitative research.

Traditional scientific research criteria – Some proponents of qualitative inquiry in its early development looked back to the traditional scientific research criteria in an effort to increase the legitimacy and credibility of qualitative evaluation. That was a somehow tentative and apologetic approach to qualitative inquiry and was indicative of a yearning to be accepted into mainstream scientific paradigm at that time. Since science emphasised objectivity as a measure against researcher bias, these researchers also stress rigorous and systematic procedures of data collection such as cross-checking and cross-validation of fieldwork, using multiple coding methods to establish validity and reliability of patterns. The purpose is both to explain and describe with maximum accuracy possible, the phenomena as they are in the real world.

Social construction and constructivist criteria – This school of thought differentiates between the social and the physical worlds, the former is viewed as a human construction, that is, it is socially, economically and politically constructed, just as human understanding and explanations of the latter are. Advocates of social construction and constructivism have developed certain concepts that are parallel to those employed in traditional scientific research. For instance, Lincoln and Guba put forward "credibility as an analog to internal validity, transferability as an analog to external validity, dependability as an analog to reliability, and confirmability as an analog to objectivity. In combination they viewed these as addressing 'trustworthiness' itself a parallel to the term '*rigor*' (Patton, 2002:267). In tandem with their social constructivist perspective, they seek to report multiple views from multiple constructions rather than a single one. Emphasis is laid on a deep understanding of specific cases within specific contexts as opposed to causal explanations and generalisations of human interactions. Patton declares that:

> "They offer perspective and encourage dialogue among perspectives rather than aiming at a singular truth and linear prediction. Social constructivist's findings are explicitly informed by praxis and reflexivity, that is, understanding how one's own experiences and background affect what one

> understands and how one acts in the world, including acts of inquiry" (Patton, 2002:267-9).

Artistic and evocative criteria – The previous two sets of criteria for judging qualitative inquiry emphasised the scientific nature of research and a mix of science and art in qualitative studies respectively. This set of criteria in turn, puts stress on 'aesthetics, creativity, interpretive validity, and expressive voice' where findings of case studies are treated as literary works and are performed. Patton writes that:

> "Artistically-oriented qualitative analysts seek to engage those receiving the work, to connect with them, move them, provoke and stimulate. Creative nonfiction and fictional forms of representation blur the boundaries between what is 'real' and what has been created to represent the essence of reality, at least as it is perceived, without a literal representation of that perceived reality" (idem: 269).

Performance of these 'creative syntheses' actually enhances understanding of the analyst's interpretation of findings while simultaneously attaching affective dimension to them. Enlightenment gained, separately or in concert with the emotional/feeling connection engendered by the analyst's expressive voice, may lead to increase interest in and consequent instrumental use.

Critical change criteria – According to Patton, the concept 'critical' generally connotes 'an interest in and commitment to social change', which makes this set of criteria biased towards social transformation. These criteria are employed by those evaluators who use Critical Theory approaches to fieldwork and their purpose is to expose social, economic and political inequalities. Rather than simply striving for understanding, they criticise the status quo, raise consciousness and endeavour to change the balance of power in favour of the less powerful or the powerless. Their perspective is Marxist to the extent that they are conscious of class differentiations in societies and they seek to make these explicit. "Critical Theory provides both philosophy and methods for approaching research and evaluations as fundamental and explicit manifestations of political praxis (connecting theory and action), and as change-oriented forms of engagement" (Patton, 2002:270). Studies in qualitative evaluation that reflect influence of Critical Theory range from research-oriented approaches whose purpose is to expose injustices to those that actually bring about change. Collaborative and participatory

approaches fall under this category because they build capacity, raise consciousness and prepare stakeholders for action in changing their conditions.

Pragmatic utilitarianism – This is the posture taken by Patton in utilization-focused evaluation. An evaluation researcher who assumes this stance is not aligned to any ideological philosophy, but is pragmatism-oriented in the sense that they would employ any model that promotes use. Patton states that the focus of pragmatic utilitarianism is on "answering concrete questions using practical methods and straightforward analysis while appreciating that those who use evaluations apply both 'truth tests' – are the finding accurate and valid? – and 'utility tests' – are the findings relevant and useful?" (Weiss and Bucuvalas [1980] as cited by Patton, 2002:271). To achieve these, an evaluator has to match questions to particular contexts and specific needs of intended users. The evaluator also has to have a wide range of skills in research methods as well as creativity.

4.10 Conclusion

Patton's contribution to the paradigms debate through his alternative approach is not altogether original. It augmented the works of other evaluation scholars, among them the prominent Egon Guba and Lee Cronbach. Nevertheless, it was a welcome voice when the qualitative paradigm needed all the reinforcements it could get for a fair recognition in its own right. Its significance lies mostly in tying arguments as presented by scholars before him together, adding to them and turning them into a composite front for a qualitative cause in a coordinated and unapologetic manner. The debate itself is presented in such a way that the weaknesses of the scientific approach are highlighted and solutions offered in the form of an alternative approach on a point-refutation basis. That has the effect of augmenting the effectiveness of the latter paradigm. The inclusion of the alternative paradigm, the paradigm of choices, serves to illustrate that even though Patton comes out forcefully for the alternative paradigm, he still acknowledges the complementary nature of the quantitative-qualitative paradigms depending on the contextual factors to which evaluation has to respond.

CHAPTER FIVE

PROFESSIONALISATION AND POLITICS OF EVALUATION PRACTICE

5.1 Introduction

In the early 1980s, the paradigms debate had raged past its apex and there was a shift away from hard-line postures explicated by adherents of quantitative and qualitative paradigms. The debates increasingly shifted towards matters surrounding the professionalisation of "evaluation" as a field. In this context, it was a critical event when the Standards of Excellence for Evaluation were published in 1981. These standards were published to regulate the manner in which evaluations were to be conducted by evaluators irrespective of their paradigmatic affiliation. The Evaluation Research Society and the Evaluation Network also merged in 1984 to form American Evaluation Association in a step that formalised a long-standing goal to achieve greater unity of evaluation research and practice. However, certain questions which cast evaluation research in dim light were still outstanding. For instance, besides a considerable growth in its body of literature and a number of full-time practitioners, evaluation research still remained largely a step child in the service of other professions without an independently acknowledged existence.

This chapter is concerned with Patton's contribution to debates surrounding the challenges of extricating evaluation research from the "shadows" of other professions and turning it into a fully fledged professional discipline recognised in its own terms. Patton suggests that in order for it to move forward and earn the respect of other professions, it had to have a vision that promotes quality which in turn require sound training in research methods and an abundance of technical expertise. While methodological competence and technical skills are critical, utility was emphasised as the vehicle to evaluation's professional acceptance. Commitment to use received a major boost with the publication of the Standards of Excellence for Evaluation which put utility at the centre of evaluation practice.

There was also a realization that evaluation needed new methods and approaches in order for it to meet its increased needs necessitated by its growth; hence propositions by evaluators like Nick Smith and Patton to look beyond their practice for new approaches and adapt them to their field to enable them to match methods to the needs of varied

disciplines that evaluation covers. Different disciplines that have similarities with evaluation are discussed in this regard.

Patton encourages evaluators to commit to the profession by taking forward what he perceives as a roadmap to the success of the profession, namely; a vision that promotes utility, quality products and processes and skilled, trained evaluators.

5.2 Patton's vision for professionalizing evaluation discipline

5.2.1 Commitment to utility

Commitment to utility means conducting evaluations that improve programmes and promote decision making about them. Patton laments the tendency among some evaluators who are hesitant when asked what they have to offer as evaluators because of his conviction that evaluation has a lot to offer. He writes:

> "We don't just offer data, although data are part of evaluation. We don't just offer reports, although reports are typically a part of evaluation. We don't just offer methods and techniques, although methods and techniques are used in what we do. What we offer are empirically-based strategies and processes for improving programs and enhancing decisions about policies and programs. That's a lot to offer! …My vision is an evaluation profession *confident* about what it has to offer, i.e., useful evaluations that make a real difference" (1990: 47).

In *The challenges of making evaluation useful*, (2005a: 71), he elucidates some of the important lessons learned about how to make evaluations useful in, namely:

- Being clear about intended uses by intended users;
- Creating and sustaining a results-oriented, reality-testing culture that supports evaluation use;
- Collaboration in deciding what outcomes to commit to and hold yourselves accountable for;
- Making measurement of outcomes and programme processes thoughtful, meaningful, timely, and credible – and integrated into the programme; and
- Using the results in a useful and transparent way, and modelling for others serious use of results.

It is apparent that Patton's vision of improving use is tied to the adoption and promotion of a utilization-focused approach to evaluation (which happens to be his

'invention') because the above tenets of increasing evaluation use are actually what informs utilization-focused evaluation. Being clear about intended uses by intended users is the first step of a utilization-focused evaluation process. He shows disposition towards formative evaluation as a means to improvement and learning because it allows for collaboration with stakeholders. In reference to it, he writes that: "a second type of evaluation is also very important, what we call "formative evaluation" or "evaluation for improvement" –an approach to evaluation that emphasizes learning, improvement, and identification of weaknesses, especially from the perspective of project beneficiaries. In such an approach to evaluation, the staff, beneficiaries, and evaluator work together collaboratively to learn how to be more effective and to change things that are not working effectively" (Patton, 2005: 69). Formative evaluation is usually undertaken internally for improvement purposes or to allow projects to develop before they are subjected to summative evaluation. Patton also believes that for evaluations to be successful, those involved in them should be willing to engage in 'reality-testing', that is, 'willing to look at what is really going on and what is actually happening, not just what we hope is happening'(idem: 70). This implies that stakeholders should be prepared to accept if things do not turn out the way they anticipated and learn from them.

5.2.2 Guidance on enhancement of quality and credibility in qualitative analysis

Quality and credibility are interrelated to the extent that the former is a prerequisite for the latter and the latter influences perceptions about the former. Issues of quality and credibility cannot be understood properly if they are divorced from the audience and purposes of research. Different audiences have different perceptions of quality and credibility just as different purposes necessitate different approaches to inquiries. Therefore, it is only logical that these differences are recognised through differences in criteria. Patton states that credibility in qualitative research is dependent on three related elements, namely:

- Rigorous techniques and methods for gathering high-quality data that are carefully analysed, with attention to issues of validity, reliability and triangulation.
- The credibility of the researcher, which is dependent on training, experience, track record, status, and presentation of self, and

- Philosophical belief in the value of qualitative inquiry, that is, a fundamental appreciation of naturalistic inquiry, qualitative methods, inductive analysis, purposeful sampling and holistic thinking (1999: 1190).

Patton (1999) states that there has been questions regarding the nature of analysis in qualitative research, hence his decision to give some guidance on it. In contrast to rules and formulae followed by statistical analysis, qualitative data analysis is a creative process and therefore, dependent on the capabilities of an analyst, both in terms of critical insights and conceptual capabilities. A qualitative analyst should be well versed in recognising patterns in the data before them and be prepared to recognise them because 'chance favours the prepared mind'. Creativity needs not be an antithesis of rigour and validity, but an additional requirement because attention to detail and systematic analysis and reporting – which imply rigour - are prerequisites for credibility in qualitative research. Qualitative analysis is also not averse to causality, but recognises its limitations. Patton, 1980 writes:

> "That's all right! When careful study of the data gives rise to ideas about causal linkages, there is no reason to deny decision makers and information users the benefits of those insights simply because they cannot be proven. What is important is that such statements be clearly qualified as what they are – speculation, conjecture, and hypothesizing" (1980: 324).

Patton highlighted the following issues with regard to the promotion of quality in qualitative studies:

Integrity in analysis – Having described patterns and linkages, the qualitative researcher looks out for different or rival explanations and organise data in ways that may result in competing explanations. They then check if those explanations can be supported by data. The constructivist interpretivist tradition is predicated on the assumption that there is more than a single reality and searching for rival explanations is part of that quest to find data that supports multiple realities, not to disprove any single explanation. In the light of massive data which is common to qualitative data collection, the researcher has to look for patterns amongst data by looking at alternative classification systems, themes and so forth as a way of 'testing' competing explanations. That lends credibility to data since it is demonstrated that intellectual rigour has been applied.

Negative cases – Part of testing patterns and trends that have been detected is identification of negative cases, that is, cases that do not fit the observable patterns. These enhance understanding of patterns by pinpointing exceptions to 'rules' or broadening them or disproving them altogether. In certain rehabilitation programmes, it may be helpful to consider and examine reactions of those who have dropped out if there are any. Dropouts may be statistically insignificant in certain instances, which is positive in terms of overall response to the programme and possible intended uses, but what is more important in qualitative research terms is the understanding their reactions elicit about a specific programmes for improvement.

Triangulation – The rationale behind triangulation is the conviction that "no single method ever adequately solves the problem of rival explanations. Because each method reveals different aspects of empirical reality, multiple methods of data collection and analysis provide more grist for the research mill" (Patton, 1999: 1192). The fact that triangulation is ideal for establishing the accuracy of findings is self-evident, but it is just as apparent that it is expensive and time consuming to apply numerous methods in a single evaluation. There are four types of triangulation in line with its areas of emphasis, but they are all aimed at verifying and validating evidence that have been gathered.

- *Methods triangulation* – This kind of triangulation often compares data collected using some types of quantitative methods with data collected employing certain kinds of qualitative methods. Since methods from the two paradigms usually answer different research questions, qualitative and quantitative methods are ideally used in a complementary manner to provide a single integrated approach to any given situation that is amenable to their joint use. Patton mentions that many researchers are not equally adept in both types of data and that procedures for merging these data are not yet in place. Therefore, investigators are inclined to relegate one type of data to the background depending on their paradigmatic socialisation. Conflicts may arise between the two types of data and they may sometimes be received in varying degrees of credibility even though they are in a single evaluation. Patton suggests that methods triangulation is a form of comparative analysis, which places it under "comparative research". He writes that according to Fielding (1986: 13), "comparative research" often involves different optional measures

of the 'same' concept, and it is an acknowledgement of numerous problems of 'translation' that it is conventional to treat each such measure as a separate variable. This does not defeat comparison, but can strengthen its reliability". Under such circumstances, it is advisable to focus on areas of convergence instead of emphasising divergence and choice.

- *Triangulation of qualitative data sources* – This is comparing and cross-checking the consistency of information derived from different times and by different means in qualitative methods. "It means (1) comparing observational data with interview data; (2) comparing what people say in public with what people say in private; (3) checking for the consistency of what people say about the same thing over time; and (4) comparing the perspectives of people from different points of view" (Patton, 1999: 1195). One may validate information derived from observations or interviews by going through documents and other written evidence and vice versa. It is expected that interview and observational data will be different; the point is to find out reasons for disparities.
- *Triangulation through multiple analysts* – Sometimes more than one analyst or observer are used in one evaluation. The purpose is to guard against or reduce bias potential of a single person doing data collection by using several observers or interviewers. This increases the validity and credibility of data gathered. The same process can be extended to data analysis, employing more than one analyst to independently analyse data from the same set of qualitative data. That counteracts bias in analysis and interpretation. One way of analytical triangulation is to involve participants in the study to review analysis. Participants check descriptions and analyses to see if they associate with them or not and the extent to which they associate with them will determine credibility of analysis.
- *Theory triangulation* – Different theoretical perspectives of the same data are employed in this type of triangulation. The purpose is to bring various intellectual and disciplinary traditions to bear on the data in order to understand how different theoretical assumptions affect findings. For example, "one might examine interviews with clients from different psychological perspectives: psychoanalytic, Gestalt, Adlerian, rational-emotive, and/or behavioural frameworks. Observations of a group, community, or organization can be examined from a Marxian or Weberian perspective, a conflict or functionalist point of view" (Patton, 1999: 1196).

Patton's vision to turn evaluation into a profession involves a drive for commitment to utility as well as guidance on the improvement of quality and credibility of qualitative analysis. Evaluators are persuaded to be forthright about what their field entails, emphasising the usefulness of evaluation. Patton discusses the challenges to utility and presents suggestions to overcome them. Quality and credibility are crucial to any inquiry and this acknowledged when Patton offers guidance on ways to enhance quality and credibility of qualitative analysis. They include methods and techniques for rigorous analysis. Patton's contribution to improvement of quality and credibility of qualitative analysis assumes special significance in the face of the so-called paradigms debate whereby qualitative studies are presumed to be less rigorous and less credible by exponent s of the quantitative paradigm.

5.3 Credibility of the researcher

This section addresses the extent to which the researcher's credibility can affect the manner in which findings are received. Researcher credibility is particularly important:

> "Because the researcher is the instrument in qualitative inquiry, a qualitative report must include information about the researcher. What experience, training, and perspective does the researcher bring to the field? What personal connections does the researcher bring have to the people, program, or topic studied?....Who funded the study and under what arrangements with the researcher? What prior knowledge did the researcher bring to the topic and the study site?"(Patton, 1999: 1198).

While there is no definite list of questions to be addressed about the researcher, it is expected that qualitative researchers should furnish any personal and professional information that may have or might have had a bearing on data collection, analysis and interpretation in the minds of users of findings. Two issues are central in determining researcher credibility, namely: researcher training, experience and preparation and researcher/evaluator and research/evaluation effects.

5.3.1 Evaluator training, experience and preparation

Patton's point of departure is that having functional senses does not make one a skilled observer. This is more so because "what people "see" is highly dependent on their interests, biases, and backgrounds. Our culture tells us what to see; our early childhood socialization instructs us in how to look at the world; and our value systems tell us how

to interpret what passes before our eyes" (idem: 1199). In the context of this selective nature of human observation, it is imperative that researchers be trained and be prepared. Patton suggests that they may be trained in descriptive writing, disciplined recording of field notes, how to separate detail from useless trivia and in the use of rigorous methods to validate data. Preparation is as pivotal as is training in making observations. Patton cites a case whereby researchers missed critical information in a test of their observational accuracy or a cynical attack at observational data by proponents of the scientific approach simply because they were not in professional observation mode, and "in the fields of observation, chance favours the prepared mind" (idem: 1200). They were not informed that their observations were being scrutinised and therefore not prepared. Preparation has mental, physical, intellectual and psychological dimensions and for the researcher to make accurate observations, these dimensions should have been prepared. Lastly, experience is a necessary ingredient for a researcher/evaluator recipe. It is Patton's contention that 'being trained as a social scientist or researcher, or having an advanced degree, does not make one qualified, competent, or prepared to engage in the specialized craft of program evaluation'(idem: 1201). One needs to get into the field to test and adjust academic knowledge to the realities of the outside world, learn and refine one's skills.

5.3.2 Evaluator and evaluation effects

Patton states that an evaluator or the fact that there is an evaluation being conducted can affect findings in four ways. They are:

(1) reactions of programme participants and staff to the presence of an evaluator;
(2) changes in the evaluator (the measuring instrument) during the course of the evaluation, that is, instrumentation effects;
(3) predispositions or biases of the evaluator; and
(4) evaluator competence (1980: 333-334).

The first effect is concerned with the claim that the fact that there is an evaluation underway may affect the performance of participants and staff positively or negatively. On the upside, they may be motivated to 'show-off' and act in an exemplary manner (the Halo-effect). On the downside, the presence of an evaluative authority may create so much tension that they underperform. Both over performance and underperformance in such cases are not true reflections of the daily operations of programmes observed, they are actually distortions triggered by the presence of researchers. The evaluator

must therefore, be on the lookout for such reactions and take them into account in his/her descriptions and attempt to make adjustments as necessary. It is a rather tricky call to make. But that can be mediated by observers staying long enough in the settings being observed in order to in order for people to get used to them and react normally.
The second effect has to do with the possibility of a change of the evaluator during the course of an evaluation. There are different reasons for such changes – it may be that the researcher has 'gone native' by taking on the culture being studied or that during his/her observations, the researcher realises that she or members of her family have the same affliction as the "objects" of her study. There are numerous reasons that may hamper an evaluator from exercising sensitivity and the consequent recall, besides political duress.

Thirdly, evaluator biases have effects on their roles in data collection and interpretation. While various efforts are done to ensure the validity of data, it is sometimes difficult to uproot predispositions ingrained by socialisation from the early years to foster impartiality. The final effects concern the competence of the evaluator. Competence includes a demonstration of capability to use verification and validation procedures, not raising expectations of stakeholders through overpromising and good performance. The researcher should strive to build a record of fairness and responsibility.

In the 1984 review of *Beyond method: Strategies for social research* edited by Gareth Morgan, Patton introduces us to a more radical perspective against the dearly held traditional scientific notion of objectivity as a means to credibility. Morgan presents "research as engagement", that is, a researcher is a person engaged in a process of creating knowledge within a specific framework. Morgan (1983:13) as cited in Patton (1984) states that science is a ""process of interaction, or better still, of engagement" such that the results of science are "as much a product of this interaction and the protocol and technique through which it is operationalized as [they are] of the object itself" (1984: 70). Possibilities of neutrality or objectivity are questioned on the grounds that there is a relationship between the researcher/evaluator and the researched/evaluated. These relationships are unique in as far as people observed, researchers, purposes of research, values and socio-eco-political environments are concerned. On account of these factors, he argues that there can be no universal criteria for judging credibility, but rather, attempts should be made to link methods to politics, values, assumptions and contextual variations. This effectively means looking beyond

methods and skills and preparing for situational factors and perspectives. This needs 'engagement', and the evaluator cannot be always correct in all 'engagements' in all contexts throughout her/his career. As highly trained individuals in rational thinking, researchers should ideally acknowledge that and change perspectives accordingly in order to create relevant and reliable knowledge. The shift in perspective also means a shift in terms of perceptions of what is objective and what is not. Judgement criteria should also be flexible enough to acknowledge various perspectives or they become impertinent to people concerned.

The credibility of a researcher plays a pivotal role in the credibility of an inquiry, particularly in qualitative research where the researcher is also an instrument. Therefore researchers have to include information on their backgrounds as it may impact on their interpretation of data. Researchers are cautioned against pitfalls that may result as a consequence of lack of training, inexperience and insufficient preparation. For that reason, Patton makes a call to evaluators to be trained in the processes of qualitative inquiry and to be prepared as much as possible in order to offset undesirable outcomes such as evaluator and evaluation effects.

5.4 Philosophical beliefs

Patton ascribes the controversy about qualitative methods to the debate in science over how best to study and understand the world. The quantitative-qualitative paradigms debate or what is sometimes referred to as logical positivism versus phenomenology stems from their philosophical differences about the nature of reality. Although the qualitative paradigm has fought its way to recognition as a credible approach, there are some concerns, which are addressed below.

Methodological respectability – Patton, 1999, points out that there was never any doubt about the appropriateness of quantitative methods, even though they too have their own weaknesses. Nonetheless, he is pleased about the ascendance of qualitative methods to the same degree of respectability, especially in evaluation research. There is now acknowledgement that researchers have to be knowledgeable about a variety of methods in order to respond effectively to specific stakeholder needs and situations. The drawback, however, is that the demands for its realisation are so many for practical realities of evaluation: for instance, narrow methodological training of evaluators, resources limitations, political interference and so forth.

Beyond the numbers game – Thomas Kuhn observed that in science 'the most held values concern predictions' and that quantitative predictions are held in higher esteem than qualitative ones. Patton, on the other hand observes that although this bias has receded, it still prevails among certain professionals in social sciences. The best way to address it is to understand qualitative methods and communicate their strength effectively, pointing out questions to which they are appropriate.

5.5 Case teaching as a tool for evaluator professional development.

Case teaching and evaluation is a call to plug the gap in evaluation training and to elevate it to the same rank in other professions like medicine, law, business and so forth where case teaching is part and parcel of professional development. Patton and Patrizi assert that "experience and reflection on experience are the best teachers, yet rich cases that capture experience for reflection have largely eluded professional training in evaluation. The absence of high-quality, readily available teaching cases has been an important gap in the field" (2005b: 5). The authors argue that traditional classroom teaching can ground students/trainees in basic knowledge of the field but it does not equip them with 'astute situational analysis, critical thinking and creativity' that are often needed in practice. Therefore, the best way to teach them these requirements is case teaching. They declare that:

> "All too often, evaluation designs, assumed to be unimpeachable, sink because of poor communication among the participants, misunderstandings, competing theories of program design, and competing methodological approaches to the evaluation itself. It can take years to hone the skills to prevent or negotiate these problems, but good case material can accelerate learning" (idem: 6).

Case teaching provides opportunities for learning from real practice examples without any pressure, they allow learners the benefits of inputs from group members on how to solve problems and resolve conflicts. Through group interaction facilitated by a trainer, learners can acquire skills applicable to different situations, which could have taken years of experience to acquire. The authors highlight several areas in which case studies can be useful, they are:

Setting goals – This is one of the first decisions evaluators have to make, to define the purpose of evaluation, how information will be used, who will use it and so forth.

Answers to these questions are crucial in giving evaluation direction and ensuring that it accomplishes what it is supposed to achieve. Novice evaluators can be spared embarrassment of learning from botched evaluations with the help of real cases that enable them to experience varied problems related to goal setting.

Diagnosis – This is critical for the success of evaluation. Evaluators have to move swiftly within temporal limitations to get to grips with organisational dynamics and discern how they affect evaluation. "A well-received evaluation is usually informed by a deep understanding of the client's needs and organizational culture. The context influences the nature of the discussion and, often, the overall framework for questions and analysis" (idem: 8). Readily available cases provide information that gives guidance on how to go about scrutinizing organisational dynamics such as its history, values, culture and so forth that may have a bearing on analysis and interpretation of results.

Managing conflict – Programmes and projects usually have several stakeholders from different backgrounds and persuasions. Conflicts are anticipated in such compositions, and a challenge is how to spot and manage them instead of futile attempts to avoid them. Suppressed conflicts tend to re-emerge, usually more grievously than before. To resolve them requires strong interpersonal skills, communication and conflict management skills. These are not readily available to evaluators who are comfortable with their methods, descriptions and figures. Case studies afford learners opportunities to practice management of such situations.

Recognising high-payoff information – Information collected through qualitative methods is usually massive, but not all of it is important. "The 20-80 rule states that , in general, 20 percent of the facts account for 80 percent of what is worth knowing" (2005: 8). The challenge then, becomes one of recognising those facts that are high in payoff and information and separate them from trivia. In the light of limited resources and time constraints, a researcher cannot look at everything in depth, and therefore, they have to master the art of finding important facts. Real case studies give learners and novice evaluators opportunities to develop the skills needed to recognise them.

Managing change – This is more area where acquaintance with cases can help students. The authors assert that there is an assumption in traditional evaluation

methods that there is predictability for programming over a period of time, and when changes occur, they do so in a linear manner. They caution that in contrast to this widely held assumption, "programs are not static, and many focus on systems change or operate under conditions of emergence. Quite often, evaluators are placed in situations where program staff are legitimately unable to predict how their theory ought to unfold, particularly when the technology is unknown or when evaluation is helping to facilitate the process of discovery and preliminary intervention trial requires rapid testing, feedback and redesign" (idem: 8). Some programmes occur in fluid political environments that are constantly changing. It is during such times that lessons learned through real case studies stand learners in good stead.

Working collaboratively – Discussions of case studies are replications of real life encounters of evaluators in the field of practice. If groups are arranged in such a manner that they comprise people from diverse backgrounds and programmes, participants stand to benefit from a wealth of perspectives on how to address issues evaluators confront in their line of work as well as benefiting from solutions offered in case studies.

The significance of Patton's insistence on utility lies in his recognition that the value of evaluation practice depends on its use. Use is connected to quality and credibility in the sense that evaluation funders, policy makers, programme managers and staff are not likely to utilise evaluation recommendations if they are not logically derived from data or when they perceive evaluation to have been poorly conducted. Issues of quality and credibility have been raised by numerous scholars such as Weiss as critical to acceptability of evaluation findings. Patton's contribution with regards to the professional development of evaluation with reference to these issues is instructive. He makes a conscious decision to approach them from an utilisation perspective; arguing that they are not ends in themselves but vehicles to use. Use is a precondition for improvement of human conditions. Quality and credibility are also related to socialisation to the extent that evaluation scholars are divided on them based on their socialisation, hence the inclusion of philosophical postures of evaluators. Patton's vision of evaluation profession is important because it offers an integrated approach to resolve problems faced by the profession, with use as the priority. This is substantial contribution towards promotion of use; building on the foundation laid by scholars such as Robert Stake in his responsive evaluation.

5.6 Barriers and challenges to the professional development of evaluation

Patton is aware that there are many challenges to the realisation of his vision for evaluation. In a further effort to professionalise evaluation, he identifies some of the barriers which have to be surmounted in order to achieve his vision. These barriers and challenges encompass real and imagined concerns of stakeholders, their education about evaluation and guidance to evaluators. Most importantly, Patton does not only raise problems, he also offers solutions to them. This is consistent with his overall vision of increasing utility of evaluations. These are addressed below:

- Overcome the fears of evaluation – as alluded to in the evaluator and evaluation effects, the presence of a person with authority or mandate to pass judgement induces anxiety among programme staff. It may be helpful to deal with these fears from the onset by encouraging intended users to voice them and allay them where possible.
- Avoid promising more than what possibly can be done in an effort to build support for evaluation – evaluation has an important contribution to make in the improvement and smooth operation of programmes, but it cannot solve all organisational or government problems. Ills of mismanagement, inefficiency, political interference in projects/programmes and so forth should not be expected to disappear after evaluation, and evaluators must desist from making unrealistic promises they cannot fulfil in their advocacy for their profession, that is counterproductive.
- Avoid equating evaluation with accountability – there is a great demand for accountability from the public services and programme directors and evaluation is sometimes perceived as enforcing that accountability. To a certain extent, evaluation can foster accountability in as far as it exposes weaknesses in institutions, but evaluators do not have the mandate to enforce evaluation findings. Cronbach and Associates cautioned: "A demand for accountability is a sign of pathology in the political system. Accountability emphasizes looking back in order to assign praise or blame; evaluation is better used to understand events and processes for the sake of guiding future activities" (1980: 4). The implication is that accountability is backward looking and judgement-oriented. While evaluation has judgement aspect to it (summative evaluation), it is more than that to the extent that it is forward looking and has influence on future

developments where it was conducted. It can also be generalized or extrapolated to other environments with similar characteristics. Patton pushes the point of the inferiority of accountability to evaluation further by asserting that 'accountability is the *lowest* level of evaluation. It focuses on the *minimum* acceptable performance' (Patton, 1990: 50). Even then the purpose is to control, rather than to achieve excellence and quality development. Therefore, it is necessary to differentiate between the two.

- Train users – 'unsophisticated' evaluation users have been identified as a barrier to evaluation use. Patton suggests that users should be guided to develop an understanding for and appreciation of evaluation. This is the burden evaluators should bear to improve utilisation of evaluation and demand for its professional services.
- Separate personnel from programme evaluation – Patton makes this call after a realisation that some evaluators and stakeholders confuse the two types of evaluation, particularly with regard to how performance indicators are linked to personnel decisions. He argues that these are different and that linkages between them should be considered carefully.
- Integrate evaluation into larger management and development processes – it is common practice in organisations or in the public sector to put evaluation at the end of planning processes, after all decisions all decisions have been made and that is how evaluation's place in the hierarchy of decision making is perceived. That has negative influence on perceptions about the importance of evaluation. Patton contends that 'useful and meaningful evaluation must be built into management and programme development processes at every step along the way – from the beginning' (idem: 50).
- Evaluation requires resources, both time and money – A decision to mandate evaluation should be followed by necessary resources to conduct it. These include financial backing and enough time to complete the study, analyse and interpret findings.
- Over-politicization of evaluation will destroy its credibility – evaluation, by the virtue of its provision of knowledge, is a political. There is a dictum that 'knowledge is power', therefore, evaluators are inherent actors in this political process. "The challenge, then, is not to remove evaluation from politics, or to remove politics from evaluation. That is neither possible nor desirable. The challenge is to be politically sophisticated in dealing with politics so as to

enhance use without undermining credibility. It is a challenge of nuances – and not one without risks" (idem: 51).

These barriers represent a synopsis of the points he raised in his vision. They are either presented as considerations to be made in preparation for evaluations or as issues to be avoided in order to maximise evaluation utilisation.

5.7 Comments on Standards of Excellence for Evaluation.

The Joint Committee on Standards for Educational Evaluations published the Standards of Excellence for Evaluations of Educational programs, Projects, and Materials in 1981. These standards had and still have great significance for evaluation practice in general and the qualitative paradigm in particular because they stipulate in no uncertain terms that "evaluations ought to be useful"(Patton, 1986:24). Before the standards, a dominant practice by the scientific paradigm was for evaluators to design evaluations, collect data, analyse, interpret and publish findings without any due consideration to usefulness of their enterprise. There was no difference between basic social research and applied social research that evaluation is part of. Patton writes:

> "Before 1978, the criteria for judging evaluation research could be scarcely differentiated fro criteria for judging basic research in the traditional social and behavioural sciences. Technical quality and methodological rigor were the primary concerns of researchers. Use was not an issue. Method decisions dominated the evaluation decision-makers process. Methodological rigor meant experimental designs, quantitative data, and detailed statistical analysis" (1986: 24).

Perhaps, the greatest beneficiary of the standards was Patton and his approach to evaluation, utilization-focused evaluation because of its emphasis that evaluations should be judged by 'their utility and actual use'. It is also worth noting that the year of change (1978) corresponds with the year of publication of the first edition of *Utilization-Focused Evaluation.* Whether the date chosen is a coincidence or an emphasis on his contribution towards that change of attitude and due recognition of qualitative data is not made explicit. But the general tone of his comments is tilted towards the latter – the standards were a 'major step in increasing evaluators' commitment to evaluation use', which is the theme running throughout his book. He quotes with gratification the summary of standards by the Chair of the Committee, Dan Stufflebeam before their publication that:

> "The standards that will be published essentially call for evaluations that have four features. These are *utility*, *feasibility*, *propriety* and *accuracy*. I think it is interesting that the Joint Committee decided on that particular order. Their rationale is that evaluation should not be done at all if there is no prospect of it being useful to some audience. Second, it should not be done if it is not feasible to conduct it in political terms, or practicality terms, or cost effectiveness terms. Third, they think it should not be done if we cannot demonstrate that it will be conducted fairly and ethically. Finally, if we can demonstrate that an evaluation will have utility, will be feasible and will be proper in its conduct, then they said we should turn to the difficult matters of the technical adequacy of the evaluation, and they have included an extensive set of standards in this area" (Stufflebeam [1980] as cited in Patton, 1986: 25).

For Patton, the shift from the traditional posture of scientific research's emphasis on aloofness and technical quality to a call for taking responsibility for use was 'a scientific revolution'.

5.8 The relationship between evaluation and other disciplines

Journalism

Metaphors for Evaluation: Sources of New Methods (1982) is a review of the book of the same title edited by Nick L. Smith. Patton says that the book is a product of a Research on Evaluation project funded by the National Institute of Education. In it, Smith states that there is a need for new methods in evaluation and that evaluators are searching for new approaches to evaluation. For Smith, the best way of meeting these needs is to look outside the field of evaluation and use of other fields '*as metaphors for evaluation'*. The goal of improvement of methods is deemed to be "optimum matches among methodology, the purpose of the individual study, and the nature of the phenomena being studied" (Smith [1981] as cited in Patton, 1982: 48). Patton declares that the book is to a large extent based on personal interests and professional backgrounds of the writers and readers because the 'authors have necessarily selected and pursued a few themes in each field that were of personal interest and seemed potentially useful in illuminating evaluation methods' (idem: 49). Nonetheless, this does not detract from the importance of metaphors. They highlight a critical aspect of teaching and training in evaluation; that of using metaphors and analogies that are comprehensible to trainees to elucidate some complex issues in evaluation. The fact that journalists' stories sell millions if not billions of news copies on a daily basis

makes journalistic metaphors more important to consider in order to boost the fledgling profession. The readers' understanding of the metaphors depends largely on their grasp of conventional evaluation practice, their acquaintance with the field from which the metaphor is taken and the extent of their encounters with the problems addressed by the metaphors. Evaluation has a lot to gain from metaphors not only on the methodological front, communication as well. Smith states that:

> "The use of metaphors may therefore not only provide use with a synthetic means of developing new evaluation methodologies…, also enable us to communicate better the nature of these new methods to evaluation practitioners and theorists"(1981: 76).

A few metaphors which are thought to be 'meaningful and relevant' have been selected to demonstrate the linkages between common evaluation problems and other fields and how the solutions from those fields might contribute towards solving evaluation problems. One such example from investigative journalism is by Egon Guba as cited in Patton: "One particular relevant concern is how journalists deal with the problem of objectivity. Investigative journalists have eschewed claims of being objective in favour of the "fairness" criterion. Attention to fairness assumes multiple realities and multiple perspectives, then directs one to strive for *balance* rather than truth" (Patton, 1982: 50). There are other parallels such as the manner in which investigative journalists trace their sources of information, how they give the subjects of investigation opportunities to defend themselves, the criteria used to decide if the story is worth writing and how they identify the audience of their stories. Although Patton does not elaborate on them, the issues related to sources of information are germane to evaluation – it relates to sampling, purposeful sampling in qualitative evaluation. Giving subjects a chance to defend themselves is relevant to phenomenological emphasis on respondents' perspective, telling it in their own terms and multiple perspectives. It would be informative to learn how investigative journalists gain trust of their sources to get information. Their criteria for deciding if the story is worth writing are pertinent to separating high-quality information from trivia in qualitative data collection.

Another metaphor which attracted Patton's attention is from Geraldine Ferguson on architecture. "The metaphor is that architecture is the development of a physical gestalt. This gestalt is a result of the analysis of the whole situation in which the solution can only be judged relative to its fit measured against context…The metaphor for

architecture being the development of a physical gestalt concerns itself with the resulting entity as a whole, made of bits and pieces, pipes, halls, concrete, functional parameters, expectations; but nevertheless, and in whatever form, a whole which exists in a context, in time and space" (Smith [1981] as cited in Patton, 1982: 50). The above extract is not concerned much about the style of architecture, but with the scope and context in which judgement is passed. Similarly, when it comes to evaluation, it would not be more about the type of method and its rigour, but about looking at evaluation holistically and the contextual factors affecting it – including participation by intended users. This demonstrates the interface between evaluation and journalism to the extent that they concerned with wholes not components in their analysis. This is something evaluation can gain by looking at investigative journalism in terms of methods and communication that invokes interest from targeted groups.

This review is significant because it highlights the shared concern of Smith and Patton about the need to increase the range of methods in evaluation and one way of doing so is to learn from other professions. Although Patton is critical of the metaphors chosen by Smith that they suit evaluators familiar with the professions from which they are taken, the two authors agree on the potential benefits of metaphors to evaluation. This need became more pronounced because evaluation was becoming more widespread in the 1980s as a result of the interlude in paradigms debate. Evaluation needed more methods to in order to be relevant to different professions. Patton responded to this need by suggesting other disciplines from which evaluation can learn. They follow in the subsections below.

Agricultural extension

Patton (1983), asserts that 'the process of making evaluation useful involves basic Extension process' and that by committing evaluators to use, evaluation standards puts responsibility on them to *extend* their findings. For him, evaluation standards are 'virtually identical' to the fundamental principles of effective (Agricultural) Extension practice. He writes:

"Evaluation can be seen as a specialized application of more general Extension principles and methods because both Extension and evaluation involve making research knowledge understandable, packaging information for decision making, educating information users, and encouraging people to act on the basis of knowledge"(Patton, 1983: 14).

The two professions – Evaluation and Extension - share a drive to pass useful information to the people for purpose of improvement. On the one hand, Extension provides information that improves agricultural productivity, nutrition and the concomitant improvement in the quality of life of communities. On the other hand, evaluation provides information for improvement of programmes, personnel effectiveness and accountability assurance. They are both aimed at alleviating human need. Patton argues that effective Extension and effective evaluation involve 'identical' principles and processes. They are addressed below.

Step 1 – The first step of both processes is identification of people to be served by both evaluation and Extension programmes. Extension programmes are not all-encompassing; they serve a certain target group. The same pertains to evaluation; one evaluation cannot serve the needs of all people.

Step 2 – Extension programme development and evaluation designs stress the importance of needs assessment, whereby practitioners attempt to find out what clients or decision makers need. This is crucial for the success of both disciplines because they may end up developing programmes or designing evaluations that are doomed to fail due to unused results if they produce what users do not want. This is likely to happen if they do not identify their needs.

Step 3 – Information disseminated by both Extension services and evaluation studies is research based. Regardless of types of studies conducted, they are both rigorous processes of research that strive to provide valid information for decision making and change.

Step 4 – This step concerns the manner in which information is delivered to the people targeted. Users of evaluation information need understandable information and preferably in their first language to apply it to their programmes. The same goes for Extension as Patton indicates that "No idea is more fundamental to Extension than the idea that research knowledge must be translated, packaged, made understandable, made practical and adapted to local situations" (Patton, 1983: 16).

Step 5 – Both disciplines encourage their practitioners to help information users not only to apply findings, but also to return to them after a certain period to follow up on progress made since they departed.

As a result of the similarities outlined, Patton finds it proper for Extension professionals to adopt standards of excellence for evaluation to advance and guide its practice. These are utility, feasibility, propriety and accuracy. Evaluation practitioners can also learn something by looking closely at Extension principles of effective practice.

Agricultural Extension has been practised successfully in many countries to promote food security since then. In the light of the substantial similarities between Agricultural Extension and programme evaluation both in terms of intentions (to improve human conditions) and in processes; and the success of the former in enhancing use, it is only logical that the latter as a new profession should learn something from Agricultural Extension, especially with regard to use. The fact that Agricultural Extension practitioners proactively engage with farmers to educate them about methods to improve production must have appealed to Patton, hence his call to evaluators to be accountable in making evaluation useful. Evaluators, like Agricultural Extension workers, should *extend* their services to stakeholders beyond just producing reports and recommendations to ensuring that evaluations are used. Lessons from Agricultural Extension can help to develop a culture of using research-based knowledge by people to improve their lives.

Quality assurance

In *Use as a Criterion for Quality in Evaluation* (2001)**,** Patton explains how programme evaluation and quality assurance started as narrow separate fields with each having its purposes, methods and applications. The two fields originally had different emphases because of their objects of study. Programme evaluation focused on programme processes and outcomes as opposed to individual processes and outcomes in quality assurance. Programme evaluation used goal-based judgement; it aggregates data intended for decision makers. Quality assurance, on the other hand, considers individual cases using professional judgement and is intended for clinical staff in the medical field. Over time, the two disciplines have expanded to a point of convergence. Quality assurance has shifted attention to issues of outcomes, data aggregation and

cumulative information over a period of time. Both disciplines now seek programme improvement and gathering useful information. Patton writes that:

> "Both functions have their origins in demands for accountability. Quality assurance began with a heavy emphasis on quality control but attention has shifted to concern for quality enhancement. Programme evaluation began with an emphasis on summative judgements about whether a program was effective or not but has shifted to improving program effectiveness ("formative evaluation")" (Patton, 2001: 158).

He declares that both fields need comprehensive programme information systems that are designed in such a manner that they allow for direct involvement of intended users, that are focused on critical success factors and that prioritise utility. They both employ quantitative and qualitative data, case and aggregate data. They have evolved from eulogising statistical processes and objectivity to embracing qualitative approaches to data collection. "The best systems of program evaluation and quality assurance will use shared processes and data and will be designed using basic principles of organizational development, lifelong learning, and positive reinforcement"(idem: 159).

Organisational development

Evaluation, knowledge Management, Best Practices and High Quality Lessons Learned (2001) constitutes Patton's critique and a warning against wholesale application to evaluation research of ideas as set out in other professions, in this instance, notions from organisational development. The critique is concerned with a transition from the information age to knowledge age whereby pressure is applied on evaluators to 'generate lessons learned and best practices' (Patton, 2001: 329). He identifies three phases of 'capitalist ascendancy', namely: the *agricultural era* when wealth was manifested in the ownership of land. This was followed by the *industrial age* with its emphasis on and domination of financial capital. The new millennium is said to be the *era of knowledge*, not just information, but knowledge, with emphasis on 'lessons learned' and identification of 'best practices'. It is sometimes referred to as 'intellectual capital' and knowledge management is at its core. Patton drives this point home by citing from the Executive Edge, 1998 a statement that: "Knowledge management is a process that harvests and shares the organisation's collective knowledge to achieve breakthrough results in productivity and innovation. In contrast, information management merely collects processes and condenses information. Knowledge

management is a collaborative discipline that aims to make employees smarter, more innovative, and better decision makers. (Patton, 2001: 330). He argues that the question of best practices, even though it has its origins in the corporate world, is quickly making its way into programme evaluation because clients of evaluators are often quick to emulate corporate practices. That affects what stakeholders expect from the findings, not just effective, but the best. However, Patton feels that the 'indiscriminate use' of the phrases 'lessons learned' and 'best practice' has devalued them both conceptually and pragmatically because they lack any common meaning, standard, and definition" (idem: 330 & 331). Moreover, he argues that when such terms are used, they are seldom given the context such as which they are best for, under what circumstances or the values that underpin what is best.

He writes: "In a world that values diversity, many paths exist for reaching such destinations; some may be more difficult and some more costly, but those are the criteria that take us beyond just getting there and reveal the importance of asking, "*best from whose perspective using what criteria*?" " (idem: 331). He contends that 'best practices' are useful only as 'principles to guide practice', but once they become 'highly prescriptive and specific'; they become 'bad practice of best practices'. He writes that personally "I consider the utilization-focused mantra that evaluations should be focused on "intended uses by intended users" an evaluation best practice *at the principle level.* However, identifying specific intended uses can only be undertaken in a specific context and situation" (idem: 331). In an endeavour to connect evaluation to knowledge in knowledge age, a distinction between *instrumental use* (when decisions or actions follow at least partly because of an evaluation) and *conceptual use* of findings (the use of evaluations to influence thinking and to deepen understanding by increasing knowledge).This knowledge can include generating 'lessons learned'. He further asserts programme evaluation has grown to a stage where the large volumes of evaluations offer an opportunity to synthesise findings from different studies to 'formulate generalizations about effectiveness'. "These kinds of 'lessons' constitute accumulated wisdom – principles of effectiveness of 'best practices' that can be adopted, indeed, adapted, to specific programs even entire organisations" (idem: 333).

But he highlights a positive contribution that Appreciative Inquiry can make in *Inquiry into Appreciative Evaluation* (2003)**.** The article compares the fields of programme evaluation and organisational development and what they can offer to each other. The

comparisons are preparation of the ground for a case for Appreciative Inquiry, which apparently began in the field of organisational development, to be applied to programme evaluation. Patton asserts that the fields developed together. Organisational development helps bring changes to organisations through conflict resolution, leadership development, solving communications problems, and developing human resources among others. It sometimes employs methods used in evaluation like surveys, field observations and action research. Evaluators also carry out development work and logic modelling is cited as a major example. He writes: "it appears that evaluators are engaged in a great deal of logic modelling, clarifying and building program theory, or otherwise engaging in conceptual mapping exercises to help program staff and other organisational stakeholders figure out what they are doing and why they think they are doing it. This sounds very much like program design and development. But it is also a critical evaluability exercise, building the foundation for subsequent meaningful evaluation" (2003: 86). He maintains that the processes of evaluation can encourage changes in organisations by helping those involved think empirically, which is, making decisions based on data. In this context programmes and projects are perceived as part of improving organisations through programme/project effectiveness. Patton uses the definition of Watkins and Cooperrider who defined Appreciative Inquiry as "a worldview, a paradigm of thought and understanding that holds organisations to be affirmative systems created by humankind as solutions to problems. It is a theory, a mindset, and an approach to analysis that leads to organisational learning and creativity" (Watkins and Cooperrider [2000] as cited in Patton, 2003: 88).

Interpreted in this context, Appreciative Inquiry is an approach to organisational development and programmes which comprise organisations that improves organisational effectiveness; this is where Appreciative Inquiry and programme evaluation intersect. In addition to integrating inquiry and action, Appreciative Inquiry emphasises the positive – what has worked, why it worked, thus encouraging stakeholders to be forthcoming with information. This is how it varies from other approaches; and it also makes the inquiry more suited to conflict situations because it does not interrogate participants about their failures. It facilitates collaboration among people who hold different perspectives by building on successes. Programmes interested in developing their strengths are more likely to opt for Appreciative Inquiry because of its emphasis on assets analysis. A distinction between assets analysis and needs analysis is made – "Where needs assessment suggests deficiencies to be

corrected, assets analysis calls attention to strengths that can be developed. Just as programs are often unaware of the needs of their clients, so too they are often unaware of the capabilities of their clients" (Patton, 2003: 91). Assets analysis is crucial in designing future programmes that contribute to development. Appreciative Inquiry to evaluation can help future programming instead of backward looking. A one-sided focus on the needs of clients to the exclusion of their strengths can lead to something along the lines of a deficiency approach whereby clients are viewed from a negative perspective and treated as perennially subject to correction. This is like traditional approach to education that Dewey opposed for treating pupils' minds as empty of experiences and therefore, to be filled in a unidirectional manner by teachers. Assets analysis does not replace needs analysis but, rather, it counterbalances it. Patton writes:

> "Evaluation reports, especially the recommendations section, tend to focus on program deficiencies, things that need to be changed or improved, and areas in which a program is weak. But programs also need to be told what they have going for them, strengths upon which they can build, and assets they can use for program improvement. A balanced evaluation report will point out program assets as well as program needs, things that should be changed and also things that should be maintained" (1981: 71).

An Appreciative Inquiry, though inclined towards assets analysis, has a major role to play in balancing evaluation's disposition towards needs analysis. People must be helped to help themselves and one way of doing that is to direct their attention to their strong points and build on them while not losing sight of the loopholes to be filled. In this way they can cease to be recipients of fish and become catchers of fish. Such an approach is good on building on the strengths for future changes, it emphasises both positive and negative aspects for a rounded development of individuals or programmes.

Futures studies and evaluation research

The Future and Evaluation (1988) is a reflection on the future of evaluation, a comparison between evaluation and the futures studies as well as an exploration of the possible role of the latter in evaluation in future. Patton observed that "The field of futures studies includes a broad range of techniques that make inquiries about the nature of the future. Futurists study the future in order to alter perceptions and actions about the present. Evaluators, on the other hand, study the past (what programs have already done) in order to alter perceptions and action in the present" (1988: 90). Whilst

the futures studies look into the future to change the present, evaluation looks at the past. There is considerable common ground between the two disciplines. The most evident is that both are interested in 'altering perceptions and actions in the present'. Patton claims that in making recommendations, evaluators act as futurists. He wrote; "In effect, at the point where evaluators make recommendations, they become futurists. Recommendations constitute a forecast of what will happen if certain actions are taken. Those forecasts are based upon our analysis of what has occurred in the past. The accuracy of such forecasts, as with predictions about the future, is subject to error due to changed conditions and the validity of the assumptions that are necessarily made" (idem: 90). The implication is that there is continuity between evaluation and the futures studies because where evaluation is about to end, that is, at recommendations stage, it intercepts with the futures studies. Therefore, the two disciplines can be interpreted as components of the same continuum, with the past and the future as extreme ends and overlapping at recommendations.

The futurists attempt to solve uncertainties of forecasts by developing several scenarios with the best and the worst scenarios at the opposite ends and other scenarios in the middle. The scenarios are useful in putting across the point that recommendations are based on assumptions that certain conditions will prevail, and should those conditions change, the programmes will be adjusted accordingly. Different scenarios also foster monitoring of more that one group of factors, thus, preparing practitioners of evaluation for a range of eventualities. That would enable evaluators to effect changes as circumstances dictate. Patton further observes that recommendations are of critical importance to the use of evaluation, yet they are the weakest link in the evaluation process because he claims, recommendations are usually poorly prepared. Evaluation, he proposes, can benefit from taking a leaf out of futures studies, especially the case scenarios.

The drive for professionalisation of evaluation was a positive indicator of the development of the field hence the need to widen its methodological arsenal and to close some gaps which might discredit it. Patton is very active in promoting this cause as articulated in his vision. It suggests ways to enhance quality and credibility of evaluation. He further makes suggestions on increasing evaluation's toolkit through his works which compare evaluation with other fields, using them as 'metaphors' for evaluation. His contribution to the professional training through case studies is

invaluable because it exposes learners to real world experiences and choices they can make in various circumstances.

5.9 The politics of evaluation

5.9.1 Political inherency of evaluation

Patton (1987) argues that politics is an integral part of evaluation and that evaluators should consider politics as part and parcel of their work instead of avoiding it. He asserts that the political inherency of evaluation is manifested in 'the political nature of its definitions'. One definition emphasises goals in as far as its sees evaluations as a process of determining the extent to which programme goals are attained. He argues that the process of clarifying and prioritising goals is a political one because it does not necessarily take into consideration all stakeholder perspectives on goals. The choice of stakeholders determines who will be empowered by evaluation, and therefore, that makes it a political issue. "Some evaluation researchers define evaluation as involving the employment of primarily rigorous social science methods to study programmes (Bernstein and Freeman, 1975; Rossi et al.,1979)" (Patton, 1087: 104). These are disposed towards experimental designs and quantitative methods. Much has been said about socialisation of evaluators and their narrow training. Deciding on a measure and how to measure something is not entirely a technical option, it involves evaluator and stakeholder interests, values, funders' demands and so on. These are issues of power relations.

Another definition puts emphasis on the comparative nature of evaluation 'as a reaction to the narrowness of evaluation when defined as measuring attainment of a single program's goals' (idem: 104). A political decision is made with regard to what to compare with what else. The fourth definition stresses the valuation element of evaluation, judging a programme's value. Judgement criteria are matters of great political contention among evaluators from different persuasions and backgrounds. There is a definition that focuses on 'generation of data for decision making and problem solving. Patton writes that:

> "when evaluation is defined as a problem-solving process (Gephart,1981) or as a process that provides information for decision making (Thompson,1975), some action process that goes beyond valuation is given primary emphasis in the definition. This approach involves the evaluator directly in the

> political processes of problem solving, decision making, and action taking" (1987: 105).

The last evaluation definition provides information to specific people (utilization-focused evaluation). Patton proclaims utilization-focused evaluation as explicitly political because identifying intended uses for intended users involves explicit choices upfront about whose interests will be represented in evaluation.

The fundamental issue here is to acknowledge that politics is pervasive in evaluation and the need to do something about it. Patton suggests that evaluation has a role to play in making a difference in programmes. According to him, evaluations based on less perfect empirical data are better than those based entirely on politics in informing decision making processes. The impact of evaluation can be enhanced by directing attention of stakeholders away from politics to the empirical nature of evaluation data. When their focus is on data, evaluation's credibility can be enhanced, thus increasing chances of its use.

Patton (1986) explicates the sources of the political inherency of evaluation. He argues that the very fact that evaluation involves people makes it a political process. Social scientists are people and as such, their practice involves their values, perceptions and politics that have an influence on evaluation throughout. Evaluation uses classification systems and categories to look at things. Patton states that:

> "The concepts, methods, theories, and propositions of scientific research are normative and value-laden. The way in which an evaluation problem is stated necessarily includes value orientations, specialized cognitive interests, and subjective perceptions about both the nature of social "reality" and what is important to know about that social reality"(Patton, 1986: 292).

These classification systems and categories have a direct influence on the kind of data to be collected. That is a political process. The involvement of empirical data in evaluation makes evaluation political to the extent that empirical data promotes knowledge. Furthermore, social science data exhibits patterns and probabilities and not conclusions about facts. These need to be interpreted into usable information. Interpretation is 'only partially a logical, deductive process' it is also value-laden and political, he argues. One would expect interpretation to be data-based not value-based.

Any deviation from this practice is actually bad interpretation and it is against the foundations of evaluation and its values of empiricism.

One more factor that makes evaluation political is its desire for actions and decisions. Actions and decisions are taken after consideration of multiple influences, and evaluation findings are but one among many inputs. Weighing these inputs and choosing among them is a political process. This is more reasonable The fact that evaluation involves programmes and organisations makes it political. These are decision making systems and as such there are almost always tensions between rational and political forces. The involvement of information in evaluation puts it firmly in the political arena. "information leads to knowledge; knowledge reduces uncertainty; reduction of uncertainty facilitates action; and action is necessary to accumulation of power"(idem: 293).

The fact that there is disagreement about definition of evaluation is a reflection of its varied purposes and different perspectives those purposes serve. Patton argues that these disagreements are political because different perspectives involve different values which compete for status and resources. Thus, the political inherence of evaluation is not only about funders, authorities or parties imposing their interests on the direction evaluation should take – it begins with evaluators themselves in their definitions of their field and positions they take in order to attract attention and resources.

5.9.2 Evaluation and democracy

The drive for effectiveness of evaluation in meeting its purposes has led to the development of many approaches; one of those is democratic approaches to evaluation. Several approaches that encourage collaboration with stakeholders such as empowerment evaluation, inclusive evaluation, utilization-focused evaluation among others, intimated democratic involvement in evaluation in their support for democratic principles of inclusiveness, social justice and in recognition of political foundations of evaluation. But they did not make the linkages between evaluation and democracy explicit. Patton traces the development of democratic evaluation in Europe to Barry MacDonald (1987). MacDonald argued that "'the democratic evaluator' recognises and supports value pluralism with the consequence that the evaluator should seek to represent the full range of interests in the course of designing an evaluation. In that way an evaluator can support informed citizenry, the sin qua non of strong democracy, by

acting as information broker between groups who want and need information about each other" (MacDonald [1987] as cited in Patton, 2002: 125). Information brokering involves rendering evaluation methods and techniques understandable to the general citizenry, negotiating with different groups and attempting to make evaluation findings accessible to the widest users possible.

These ideas were developed further by Saville Kushner (2000), who advocated evaluation as a 'personal expression and political action with a special obligation to be critical of those in power'. This view advances people's experiences of programmes to the centre of evaluation because they are supposed to be beneficiaries. Therefore, their experiences and perceptions should count for something; they should not be treated as intellectual imbeciles and passive recipients of charitable knowledge. People's judgements are above those of outside professional experts and evaluations should strive to capture and report their views, he contends. He calls that 'personalizing evaluation' with the intention of strengthening democracy. Kushner envisages an environment:

> "Where each social and education program can be seen as a re-affirmation of the broad social contract (that is, a re-confirmation of the basis of power, authority, social structure, etc.), each program evaluation is an opportunity to review its assumptions and consequences. This is commonly what we do at some level or another. *All programs should expose democracy and its failings; each program evaluation is an assessment of the effectiveness of democracy in tackling issues in the distribution of wealth and power and social goods.* Within the terms of evaluation agreement, taking this level of analysis into some account, that is, renewing part of the social contract, is to act more authentically; to set aside the opportunity is to act more inauthentically, that is, accept fictions" (Kushner [2000] as cited in Patton 2002: 126).

Ernest R. House and Kenneth R. Howe have made a strong case for democracy with their approach, **deliberative democratic evaluation**. They set three requirements for its operationalisation, namely; inclusion, dialogue and deliberation.

Inclusion – this requirement states that deliberative democratic evaluation should include all relevant interests. They argue that it is immoral for evaluators to conduct evaluations for the most powerful interests only or to sell them to the highest bidders for their uses. This is prejudicial against interests of the poor who cannot afford the

necessary financial incentives needed to purchase evaluator services and therefore, access to evaluations. Powerful interests can abuse outcomes through omission and commission by deleting what they do not like or accentuating what they like because their interests are not always congruent with those of the less powerful or powerless. They write:

> "The most basic tenet of democracy is that all those who legitimate, relevant interests should be included in decisions that affect those interests. When only a few decide social policy, an aristocracy, plutocracy, or technocracy exists, depending on whether talent, money, or expertise is the source of authority. We think that evaluation in democratic societies should be explicitly democratic, while recognizing that there are different conceptions of democracy and that evaluative expertise has an important role to play" (House and Howe, 2000:5).

Therefore, democratic deliberative requires that the interests of all relevant stakeholder groups be equally represented centrally in an evaluation to avoid power imbalances in them that may direct them in favour of some interests at the expense of others.

Dialogue – deliberative democratic evaluation should be dialogical. There may be some hurdles to be jumped to achieve a real dialogue because some individuals and groups may not be empowered sufficiently to determine their interests on their own. They may be misled by unscrupulous elements within the media or powerful groups whose interests might be protected by suppression or twisting of evidence from evaluations; or their interests may be innocently misrepresented. In writing about choice, the authors explicate that:

> "The real interests of an individual or group are not necessarily the same as perceived interests. Real interests might ideally be defined thus: Policy X is in A's interests if A were to experience the results of Policy X and Policy Y and would choose the results of Policy X rather than that of Policy Y. Identifying real interests is critical. Choice alone is not necessarily determinative of real interests. It must be choice exercised under the right conditions" (2000: 7).

Deliberation – evaluations should be deliberative. 'Deliberation is fundamentally a cognitive process, grounded on reasons, evidence and principles of valid argument, an important subset of which are the methodological canons of evaluation' (idem: 8).

Deliberative democratic evaluation is different from other approaches of democracy to evaluation because it does not take preferences, values and interests of citizens as given with the evaluator only maximizing and satisfying them regardless of what they are. It advocates examination of preferences through rational processes. The authors write:

> "Evaluation is a procedure for determining values, which are emergent and transformed through deliberative processes into evaluation findings. Evaluation thus serves a deliberative democracy, one in which interests and values are rationally determined; and careful discussion and determination require the expertise of evaluators, often acting as experts with special knowledge" (idem: 8).

Patton suggests that a healthy democracy is dependent on informed citizenry, and one way of informing them is to help them *weigh evidence* and to *think evaluatively*. Responding to ideas from philosopher, Hannah Arendt, who believed that 'totalitarianism is built on and sustained by deceit and thought control' and that these can be countered if people practice thinking, Patton asserts that evaluation through process use can fulfil the purpose of practicing thinking. Earlier, we mentioned that Patton defined process use 'as relating to and being indicated by individual changes in thinking and behaving that occur among those involved in evaluation as a result of the learning that occurs during the evaluation processes'. Therefore, process use of evaluations can strengthen democracy by helping people to think evaluatively, thus producing a more informed citizenry and by entrenching a more thoughtful and deliberative citizenry.

Pollitt (1999) has a slightly different perspective on democratic evaluations in so far as they are dependent on collaboration with stakeholders. In response to some sweeping statements that stakeholder involvement is fundamental to successful evaluation practice, and it argues that collaborative evaluation is not the answer to every evaluation question. The critique is centred on what Pollitt terms the three 'virtues' of the collaborative approach as propagated by its proponents. The first is the claim that collaborative evaluation is likely to lead to greater external validity because it embodies a multiplicity of perspectives as represented by various stakeholders. To this claim, Pollitt questions the feasibility of such a scenario whereby the views of a diverse group of stakeholders can be equally represented. He contends that in circumstances where the winner takes all or when there is no chance of win-win, there is no point in compromising. In such a situation collaboration can only be imposed, thus causing

hostilities among constituent stakeholders. It is further claimed that collaboration has been successful only in instances where interventions were on a small, local scale. The second purported virtue, of which Patton is a strong exponent, is that collaboration with stakeholders leads to greater utilisation of evaluation findings. On this point, Pollitt aligns himself with Weiss' assertion that use of evaluation is not really dependent on the evaluator because utilisation comes about in an incoherent and winding manner that in the end, there is generally no conclusive evidence that use is directly attributable to collaboration. Where there is evidence, it is modest. He argues that chances of this happening are probably increased if they (policymakers) have been intimately involved with the development of the evaluation and that many of Patton's examples and arguments point the same way.

The third claim is that collaborative evaluation is ethically superior because 'it is founded on an equitable and democratic acceptance of the right of every stakeholder to have a full and equal say in the conduct of an evaluation' (1999: 80). It has been mentioned already that it is not always possible to reach a consensus among stakeholders. Secondly, it is argued that a consensual majority view does not always represent justice or moral and ethical high ground. History is awash with examples where majority decisions have negatively affected minorities; therefore, majority consensus is not tantamount to ethical correctness. Furthermore, he argues, collaboration may not be the only ethical consideration in certain circumstances, and as such, it may have to be traded with other stronger ones.

5.9.2.1 Factors that affect democracy in evaluation

Transparency – a major challenge to promoting accessibility in evaluation relates to transparency, that is, ensuring that the public has full access to evaluation findings. There have been attempts, some successful in certain quarters to suppress or change evaluation findings or delay evaluation reports. This leads to demands for more transparency when taking evaluation to the people. Transparency will have more meaning if reports are understandable, relevant and usable to users and when they are published in public media.

The need for evaluation – this issue is concerned with whether people really need an evaluation or it is imposed. Patton states that in many instances he found that people had had disappointing experiences with evaluations. For example, they may had been

conducted in an abrasive and domineering manner that programme managers and staff only served to respond to questionnaires followed by harsh judgemental reports on their performance with very little or no contribution towards improvements. Such negative encounters cause them to view evaluations in pessimistic light. It is the evaluator's duty to change negative perceptions about the practice generated by bad experiences with it. He writes:

> "Part of the context for each evaluation is the historical experience of evaluation that conditions and affects new evaluation design and implementation efforts. Moreover, for our ongoing professional learning, we need to evaluate our evaluations to find out how they are actually used and become more sophisticated about and adept at doing useful evaluations. The result of this will be better delivery on positive promise inherent in the idea of taking evaluation to the people" (Patton, 2002: 131).

Useful evaluations are those that focus on issues relevant to the people and that treat them with respect in their engagements in order to make discussions meaningful.

Communicating with people – the language barrier is critical in taking evaluation to the people. Difficult academic jargon distracts from real contents of reports and drives a wedge between evaluators and users. It is imperative that reports are written in a language accessible to users. Patton recommends that evaluators should build on what people already know (from the known to the unknown), metaphors can go a long way to illuminating difficult issues.

Decentralisation and evaluation – 'decentralization creates tensions in the interface between bottom-up, participatory approaches to evaluation on the one hand and more managerial modes of evaluation- compliance, control, and auditing – on the other hand"(idem: 132). It is his view that these tensions are 'natural and inevitable' and as such more energy should be directed at managing them rather than trying to eliminate them. Evaluation can contribute towards this by creating dialogues among centre-based stakeholders and between regional and local stakeholders. This will cater for differences in expectations between affected supposed beneficiaries whose needs evaluation is meant to alleviate at local levels on the one hand and decision makers at regional or national levels on the other. These differences should be reflected in the multiplicity of designs as well as in the interpretation of findings.

The main features of democratic evaluation are its insistence on representativity of stakeholders and its quest for social justice. An evaluation that is representative of a wide range of stakeholders and their interests promotes the notion of informed citizenry. This goes a long way in bringing about social justice in communities by closing knowledge gaps between communities and among members of communities while simultaneously giving all views represented an equal chance with regards to possible advancement. Evaluation should not be divorced from social, political and economic problems people face in their communities or it will come short of its ultimate goal – social betterment.

5.9.2.2 Effects of democratic participation on methodological quality of evaluation. Concerns have been raised that participatory and collaborative approaches to evaluation (they create conditions for democratic participation) reduce its methodological quality. It has been mentioned earlier that democratic evaluation requires inclusion, dialogue and deliberation and for these to take place, an evaluator has to come down to the level of other stakeholders instead of exercising his/her methodological prowess to their exclusion. This dilutes methodological and technical quality. Patton (2002) contends that for evaluation to contribute to strengthening democracy, it has to rely on the very aspects that democratic evaluations purportedly diminish, that is, quality, validity, credibility and independence. However, he attaches a contextual twist to the matter technical standards. He asserts that "the issue is not meeting some absolute research standards of technical quality but, rather, making sure that methods and measures are *appropriate* to the validity and credibility of a particular evaluation purpose and specific intended users"(Patton, 2002: 134). He refers to the study conducted by Jennifer Greene on technical quality against user responsiveness to give impetus to illustrate his point. Greene wrote:

> "Evaluators should recognize that tension and conflict in evaluation practice are virtually inevitable, that the demands imposed by most if not all definitions of responsiveness and technical quality (not to mention feasibility and propriety) will characteristically reflect the competing politics and values of the setting"(1990: 273).

The best way to move forward is to acknowledge and address the political inherency of evaluation by describing clearly the politics and values that undergird decisions taken about designs and methods chosen. This issue involves trade-offs between truth and

utility whereby stakeholders apply 'truth tests' to decide on the reliability of evaluation and 'utility tests' to check its usefulness and relevance (these issues will be explored further in the next chapter). In an ideal world, both are indispensable part of a high-quality evaluation, but the realities of practice sometimes necessitate trade-offs between them. Several factors affect choices to be made, for example, *time*. Patton (2002) claims that in most cases timelines are so short that an evaluator has to choose between highly accurate data outside timeframes and less accurate information within timelines. There is more likelihood that the former will not be utilised because it comes after the fact than the latter which meets deadlines. His experience is that "decision makers prefer some information to no information" (2002: 135). The *context* of the study also plays a role in trade-offs made. For instance, the political environment and the extent of uncertainty that the decision makers are confronted with and the availability or scarcity of information amid that uncertainty determine its urgency. If information is scarce, new information is usually helpful, even if it is not highly accurate. Inversely, information may be needed for long-term policy planning. The information provided for that may be subject of debate s by politicians or decision makers in order to ensure that they make good policies founded on facts. The evaluator must be in a position to defend the credibility of such information and therefore, it should be accurate – truth takes precedence, but such information must also have utility. *Scepticism* about returns on investments on evaluation has effect on sponsors' choice of studies. They are averse to spending large sums of money on large-scale, elaborately designed evaluations that produce little useful results. Patton writes that Cohen and Weiss conducted a review of 20 years of policy research on race and schools (1977) and they found substantial progress research methods and techniques (bigger sample sizes, computer technology, multiple regression and analytic techniques, etc.,) but

> "After reviewing the findings of studies produced with these more rigorous methods, as well as the uses made of their findings, they concluded that 'these changes have led to more studies that disagree, to more qualified conclusions, more arguments, and more arcane reports and unintelligible results"(Patton, 2002:135-6).

Moreover, Patton claims that technical quality may not get much attention because stakeholders are not knowledgeable in them, but in the same breath, he claims that the very ignorant stakeholders/non-researchers 'know (almost intuitively) that *the methods and measurements used in any study are open to questioning and attack'*. It is

paradoxical that Patton believes that evaluation research which is supposed to be a highly cognitive exercise is amenable to basic instincts of intuition. Credibility involves more than quality, but utility of evaluation, the important factor is that an evaluator explains the imperfections that are there and the reasons for them to stakeholders so that they know the strengths and weaknesses of data.

Collaborative approaches to evaluation involve measures of democracy in varying degrees even though the linkages between evaluation and democracy are not overtly stated. That is probably because of the undesirability of politics in evaluation as demonstrated in the negativity attached to 'political use' of evaluations or their findings. There was also a time when politics was not supposed to be part of evaluation; and democracy is unequivocally a political term. The recognition and admission that politics is an unassailable part of evaluation through the writings of Patton, Weiss among others, may have contributed towards explicit links between evaluation and democracy being made. The deliberative democratic evaluation makes the most explicit connection between the two concepts and practices and it also maps out the requirements for deliberative democratic evaluations. Advocacy for democracy in evaluations is predicated on the assumption that multiple views represented by such evaluations increase their external validity as well as enhance chances of utilisation. It is also contested that accommodating views majority perspectives lend moral and ethical superiority to democratic evaluations. Factors which determine the promotion or suppression of democracy in evaluations are highlighted with a view to increase democratic participation as well as the effects of such participation. The discussion is balance with both positive and negative aspects being expounded. However, some evaluators are not impressed by collaborative approaches and their democratic practices. Pollitt disputes the claims to greater external validity and increased utilisation in collaborative evaluations on the grounds that it is possible to represent all stakeholder perspectives equally and that such discrimination may lead to hostilities. Following on Weiss, Pollitt claims that collaborative approaches may only succeed in small programmes and that utilisation is unplanned and indirect. He further questions the presumed moral and ethical superiority of consensual majority decisions and their factual correctness. Decisions of majority are not necessarily ethically appropriate or factually accurate.

5.10 Management of politics in evaluation.

Now that it is clear that politics is an integral part of evaluation, the issue becomes one of managing it, not of avoiding it because that will be a futile exercise. 'How does the evaluator take politics into account without becoming a political tool of only one partisan interest?' Patton asks. The traditional scientific perspective to minimize the impact of politics is to maintain distance from the people they study in order to retain objectivity. The collaborative human relations perspective favours interpersonal contact in an effort to build mutual understanding. The two sides fall back into their respective camps according to socialisation. Patton on the other hand, proposes that political influences can be managed by focusing attention of decision makers and other stakeholders 'on the empirical nature of evaluation' He writes:

> "The empirical basis of evaluation involves making assumptions and values explicit, testing the validity of assumptions, and carefully examining a program to find out what is actually occurring. *The integrity of evaluation depends on its empirical orientation*. Likewise, the integrity of an evaluation group process depends on helping participants adopt an empirical perspective" (Patton, 1987: 130).

Therefore, despite common evaluation restrictions like temporal and financial restraints, evaluators and stakeholders must try their utmost to discover what is really happening. Such a process takes into consideration situational variations and 'explicating the politics and values' that lay the foundation for decision making about designs and methods chosen. It appears to me that despite a great effort by Patton to explicate how to manage politics of evaluation, he cannot escape from his question: 'how to manage politics so as not to become a political tool of one partisan interest'. Utilization-focused evaluation is particularly partisan in its emphasis on intended uses for intended users and no matter how empirical it is in its conduct, it remains very partisan at another level. Empirical orientation is intrinsic to evaluation conduct, which is one level, but standing from outside of the chosen few, one cannot help realizing how exclusive and partisan it is. Perhaps that is an issue of coverage of evaluation – utilization-focused evaluation has micro-coverage as opposed to macro-coverage that House and Howe advocate in democratic evaluation. U-FE relationship to other broader evaluations is analogous to focusing on improving home relations in order to improve community relations instead of focusing on the community directly. Nonetheless that does not reduce its usefulness in its scope, the question is; will it permeate beyond its

small enclave? Questions may also be asked about democratic evaluations; are they not biting too much to chew?

In *Managing the politics of evaluation to achieve impact* (2006), Mohan and Sullivan describe the question of politics in evaluation as a difficult one considering that there is not one commonly accepted definition of politics. Moreover, available definitions seem to suggest that politics permeates every sphere of decision-making. This leads the writers of this study to ask 'should evaluators avoid politics? And that 'can evaluators conduct evaluation in a vacuum devoid of any politics and still be effective in influencing the public policy process? They seem to concur with Patton's account of political inherence of evaluation when they state that "whether by force, persuasion, intimidation, or by inspiration, "A getting B to do H" is considered central to a political event" (Mohan & Sullivan, 2006: 7). This underlines the broad nature of politics as used by different evaluators because they practice in an environment where there is interaction among divergent interest groups with unequal power, influence and authority. They assert that even though it often becomes more discernible at the report stage, the intersection of politics and evaluation does not occur only at the end, but throughout the evaluation process, from planning, implementation, dissemination and so on.

Maximising both independence and responsiveness: Independence on the one hand, usually means the evaluator keeps the distance with stakeholders in order to promote credibility of evaluation. The down side is that it involves the risk of being unresponsive to the context and the needs of clients and the associated non-use. On the other hand, responsiveness brings evaluators closer to stakeholders and their politics, which may smudge the distinction between evaluator and stakeholder roles, thus resulting in loss of credibility. They contrast with Patton on the issue of trade-offs because for them the solution to this dilemma is not to trade one for the other, but rather, to make an effort to balance the two, they argue. This balance can be achieved by employing certain strategies which include " identifying stakeholders; understanding the political context, including the relationships among key stakeholders; Consulting with and responding to the needs of sponsors and stakeholders; managing the project's scope; appreciating and responding appropriately to the timing constraints under which policymakers operate; judiciously exercising statutory authority providing access to data; using evaluation methods that produce accurate information based on a broad

range of perspectives; using professional standards to guide the evaluation work; and preparing reports that are balanced in content and tone" (idem: 12).

5.11 Ethical issues in qualitative research and evaluation

Ethical considerations have been included because they are necessary to keep political interference and self-interest in check. Evaluators, both internal and external have a vested interest in their careers and monetary gains that accrue from evaluations. The situation is more challenging to internal evaluators who also face issues of power relations. Their position in the hierarchy of power within their organisations has significant influence or lack thereof, on what is evaluated and not evaluated because it affects their independence to make decisions regarding areas/departments to be evaluated within organisations. In certain institutions evaluation units are located at such a low level that they do not have any leverage as to what is to be evaluated; they are only good as management tools to achieve their prescribed ends. They receive instructions from the employer and refusal to carry them out amounts to insubordination. In such instances, evaluators are in conflict with their professional commitment and put their integrity on the line in pursuit of career protection and concomitant financial security. The question of whether or not they have a choice to be ethical is itself a subject of ethical debate that appears to be inconclusive, is opting for righteousness that results in expulsion and a number of empty stomachs going to improve anything – or ethically correct? The answer is highly personal and open. In extreme cases, performance appraisal only happens during employees' assessment period and the process is sometimes misused by management to block employee upward mobility or to gain stranglehold on them.

The issue of ethics is not limited to predicaments of internal evaluators and performance appraisal only. Another ethical question concerns 'informed consent' whereby respondents are informed about interviews and their options. The sticking point about informed consent is that it is increasingly becoming evident that interviews can be transformative. Patton states:

> "We are becoming more and more aware, for example, that interviews can be and often are interventions. They affect people. A good interview lays open thoughts, feelings, knowledge, and experience, not only to the interviewer, but also to the interviewee" (2002: 277).

The problem then is, at commencement of the interview, neither the interviewer nor the interviewee is aware of how interviews might affect respondents to speak with authority about informed consent. It is an accepted practice in interviews to disguise locations and identities of respondents in order to establish confidentiality, which safeguards them against threats, persecution or worse afflictions. But certain groupings take pride in their identities, especially when there is something to be gained by such divulgence. Patton cites political and empowerment groups that insist on using their respondents' names, perhaps to score some political points or to empower participants by having their identities attached to their stories. This is perceived as empowerment or healing of some sort by owning one's story. An evaluator then finds themselves in an ethical predicament whether to allow respondents to have their way and a sense of empowerment and healing or disempower them in the name of privacy. Is it patronizing to register in them the importance of confidentiality lest their identities extend beyond them to their families? Which is ethical – their empowerment or condescending protection? The matter of compensation and its relationship to response rates and data quality is also raised. Compensation increases response rates and honesty of responses, thus enhancing data quality, it is claimed. The question is whether to appeal to respondents by impressing upon them their contribution to and their appreciation of knowledge creation and programme improvement or to give them financial incentives. Patton avers that " in Western capitalist societies issues of compensation are arising more and more often both because people in economically disadvantaged communities are reacting to being overstudied and undervalued, and because private sector marketing firms routinely compensate focus groups and non-profit sectors" (idem: 279). Is the trend ethical? Is it not going to reach a stage where only those who compensate the most will secure interviews?

In addition to concerns about general ethical issues in evaluation research, Patton highlights some ethical issues specific to utilization-focused evaluation. He refers to three ethical concerns that have consistently come up with regard to it, namely; " (1) limiting stakeholder involvement to primary intended users, (2) working closely with those users, and (3) judging an evaluation by the extent to which it achieves intended use by intended users" (Patton, 2008:546). He contends that the best way to manoeuvre out of ethical dilemmas is for evaluators to be 'deliberate and intentional' about their moral standing. They should explicate from the onset during negotiations about the design, whose interests are to be represented in questions to be asked and clarify issues

of accessibility to evaluation findings (who will access them). He argues that the *active* part of 'active-reactive-interactive-adaptive' slogan/process is about evaluators tabling their issues and values. In this way they are able to present their own morality and integrity while adhering to professional principles and respecting the needs and beliefs of other primary intended users. What if an evaluator's sense of morality is in conflict with those of primary intended users? Patton's response would most likely be a withdrawal from that evaluation. He indicated that in response to Ross Connor when asked, "You pick and choose clients, right?" "I affirmed I did". This brings us to the issue of "clientism" – what the client wants is ethically correct, which is in the same breadth as 'the customer is king'. Patton dismisses that on the grounds that too many abominations have been committed in the name of clientism, for instance, weapons of mass destruction, concentration camps and so forth. Ernest House once raised the question of limiting evaluation to serving clients to Patton who feels strongly the responsibility to serve his clients. House fervently argued that it was immoral to do so because it implies serving only those who can purchase evaluations. To that accusation, Patton responds thus:

> "I'm reluctant, as a white, middle-class male, to pretend to represent the interests of people of colour or society's disadvantaged. My preferred solution is to work to get participants in affected groups representing themselves as part of the evaluation process" (idem: 549).

This is indeed an eloquent and noble response in its drive for empowerment in the form of self-representation by affected segments of the population instead of being patronisingly represented by outsiders. However, this is not an ideal world – those who are in desperate need of betterment seldom constitute evaluation task groups, and if there are no advantaged and enlightened people to advocate on their behalf or to actively conscientise them, they will remain perennial objects of evaluation studies that rarely upgrade their conditions. Struggles across the globe have been led by advantaged people who used their enlightenment to conscientise their people to act to liberate themselves. Besides, a qualitative researcher/evaluator of Patton's calibre is supposed to be aware of the dictum 'immersion in the settings of units of analysis' to gain the 'emic' perspective and empathy to report phenomena 'in their own terms'. These are conveniently forgotten when it comes to utilization-focused evaluation.

5.12 Conclusion

The decade leading up to 1990 was a 'golden age' in evaluation research. It witnessed a lull in the paradigms debate which provided space for a shift in attention from quarrels to real issues that evaluators were confronted with in the field of practice, concerned with how to improve the conduct of evaluations. It was a decade during which positive developments such as The Standards of Excellence for Evaluation were published (1981) by The Joint Committee on Standards for Educational Evaluation. It also witnessed a watershed merger between the Evaluation Research Society and Evaluation Network to form the American Evaluation Association, which in turn grew in leaps and bounds to 'nearly 3000 members, half of whom identify evaluation as their primary field of professional practice' (Patton, 1990:45) by 1990. There was a corresponding growth in evaluation literature that amounted to 'hundreds of books and specialized journals p.45). It was a decade of hope indeed, but Patton envisioned an evaluation field beyond that phase to the one to which it would be recognised as a profession. In his address to the Australasian Evaluation Society in Queensland, 20 July 1989, mapped out his vision of the envisaged profession thus:

> "There is a growing interest in evaluation worldwide. I fully expect to see the formation of an International Evaluation Association in the near future. This worldwide interest in evaluation is the context of my remarks today. The challenge I address is what it means for evaluation to be a serious field of professional practice" (idem: 45).

By professionals, Patton refers to 'internal and external evaluators, full-time and part-time evaluators and everyone engaged in evaluation seriously', but he is not oblivious to the fact that not all who are involved in it has interest in its professional standing.

The evaluation field is gradually shifting away from the stance which formerly perceived politics as a threat to a stage where politics is viewed as part of the profession. While political wrangling is implicit in evaluation's definitions, some practitioners have moved on to propagate explicit political agenda for evaluation as expressed in approaches such as empowerment and deliberative democratic evaluations. However, there is a need for evaluations to be conducted properly and in spite of some disagreements concerning certain ethical issues, ethics remain a critical part of the conduct of evaluation.

CHAPTER SIX

PATTON'S INVOLVEMENT IN DEBATES IN THE AMERICAN EVALUATION ASSOCIATION

6.1 Introduction

This chapter discusses some of the debates within the American Evaluation Association in which Patton was involved. I commences with the ongoing but elucidating debate between Patton and Carol Weiss on planned, direct use and unplanned, indirect utilisation, followed by similarly illuminating exchanges between Patton and Michael Scriven on purposes of evaluation. Patton's debate with Scriven was sparked by the latter's definition of formative and summative evaluations in "Beyond Formative and Summative Evaluation" (1981). Patton felt that the definition was restrictive about evaluation purposes in the light of the growth of the profession. Their exchanges reveal glimpses of their postures on the paradigms debate. The last debate concerns exchanges between Patton and Fetterman on empowerment evaluation. These exchanges were initiated by Patton in his responses to the two authors' works. In contrast, the debates with Henry and Kirkhart were reactions to Patton's writings. Both authors are uncomfortable with the pervasiveness of use or utilisation in evaluation literature as spearheaded by utilization-focused evaluation. Henry emphasises social betterment as the ultimate purpose of evaluations while Kirkhart points out the limitations of use in covering process-based influence. The contributions of Henry and Kirkhart are examined at length to provide the background for their arguments because they are premised on their theories of social betterment and influence respectively. Finally, I have added short critiques of theory-driven, feminist and fourth generation evaluations by Patton.

6.2 The Weiss-Patton debate on the limits to evaluation use

The debate was triggered by a paper Carol Weiss presented in the American Evaluation Association's conference in 1987 where she expounded the view that many decisions are a result of long processes which take into consideration different and often conflicting needs of stakeholders in pluralistic communities. She further declares: "But in most cases that I know of, the influence of evaluation on program decisions has not noticeably increased. Even evaluators who have tried conscientiously to abide by the traditional advice have had indifferent success in making evaluation the basis of decisions" (Weiss, 1988a:173). Weiss' pronouncement about the 'indifferent success'

of evaluation in influencing decision making processes did not go down well with Patton and it was the beginning of a number of illuminating exchanges between them. Weiss put forward the following reasons for that 'indifferent success'.

Multiplicity of information sources – The traditional advice, she claims, does not take into cognisance the multiplicity of stakeholders involved in decision making and that 'decisions are the outcome of interests in contention'. A number of factors that counter the ideal are then put forward. Among them is the multiplicity of interested parties in policy determination. For instance, policy makers as politicians hold certain ideologies that underpin their approach to issues. Business introduces an economic dimension that has to do with markets and profitability, some want to advance their influence, careers and so forth. Therefore, evaluation findings are but one of the competing interests in policy formulation, and it is presumptuous to think that evaluation occupies a special place among other interests to warrant it being the sole basis for decision making. She then pronounces on several ways evaluation ideas travel in a policy making environment which are meant to support her claim that evaluation is not the sole source of information to policy makers. They serve to demonstrate the multiplicity of influences and the indirect manner in which decisions are arrived at in policy making.

Policy maker's staff – Weiss avers that policy makers rely on their staff to keep them informed about developments in order to keep them abreast of their opposition. The aides have a wide range of material to pick up information from, as opposed to a particular evaluation study. The more studies converge on findings, the more current those studies become, thus having influence on policy making decisions.

Blue-Ribbon Commissions and Panels – These are special commissions established by presidents, congress and cabinet secretaries to look into issues like law and order, state of education in the country and so forth. Members of commissions represent major sections/sectors of society like industry, labour and education. They also employ professionals to conduct research, to collect information from existing research and evaluation and to write reports. These commissions reportedly have a special appeal to policy makers because of their representativeness. It becomes easy for policy makers who receive inputs from diverse interest groups in society to reach consensus, especially because these interest groups are most likely to comprise constituencies represented by policy makers.

Mass media – The media, through reports in major newspapers assist in making research and evaluation topical in society. Apart from alerting policy makers to information they might have missed, the reports also challenge policy makers to think carefully about them knowing that they are in the public domain, and they may have to respond. "It is not just that *Time* magazine or the *Washington Post* gives a capsule summary of the evaluation; more important, policy makers know that the same story reaches all the other players in the policy game. They will be asked about it. They had better know about it. They can't sweep it under the rug" (idem: 179).

Interest groups – Many interest groups use evaluation data to advance their interests. Interest groups like unions, associations *etc.* - though not disseminating evaluation data - use them to advance their interests when they are congruent with their causes. They present different kinds of evidence available to influence decision makers. Policy makers are said to like the kinds of arguments put forward by competing interest groups because they make their job easier.

Issue networks – In highlighting the importance of issue networks, Weiss writes: "All kinds of people are members of networks – staff of executive-branch agencies, political legislators, program managers, academics, consultants, interest group representatives, members of Congress and their staffs, state executives and representatives, and "think tank" experts" (idem: 180). These members are said to have become influential in moulding policy because of the composition of these networks. They have the time to communicate and come up with representative views during the gestation periods of the policies, thus, acting as catalysts in the policy making process.

The scope of evaluation – This second limitation to evaluation use concerns evaluation coverage in relation to the range of organisational issues programme management, administrators and staff have to grapple with on a daily basis. Weiss claims that "evaluation is never comprehensive or convincing enough to supply the correct answer". There are always issues that evaluation does not address, but that decision makers have to worry about: What changes will make the program work better? How much will such changes cost? How hard will it be to implement change? Which groups will support or oppose it? What will constituents think?" (idem: 173). While these claims in support of her claim to evaluation's lack comprehensiveness can be

anticipated and therefore, answered by the utilisation-focused evaluator, especially an internal one, it is a fact that the evaluator's influence within an organisation can only go a certain distance, not throughout it. Weiss sounds comfortable with 'indifferent success' of evaluation use because she does not actually propose ways to increase it. It is not clear from Weiss why organisations or policy makers would commission and spend money on evaluation unless they need information.

Rationality of decision making – The third reason for the dearth of evaluation use is that in many cases programme decisions are not a result of a rational decision process, but of decision accretion. According to Weiss, this is when a decision is taken unconsciously through unrelated steps taken in different departments/units for different purposes within organisations only to find that they have actually shaped a course of action for a certain indirectly affected department/unit. In such instances, a decision just comes about without anyone formally making it, it happens by unconscious precedent setting. Weiss calls this decision *accretion*. She writes:

> "Small steps are taken in many different offices that close off some avenues and narrow the range of available options. By a series of disjointed and amorphous accommodations, a line of action takes shape, which only in retrospect is seen to be a decision….When this is the case, the use of evaluation evidence is difficult. If people in the organization are not making formal choices, they do not go out and seek information – nor are they particularly responsive to information coming in. They act on the basis of what they already know – their current stock of knowledge" (1988a: 173-4).

Weiss states that evaluators conduct their studies with the intention of assisting decision makers to make informed decisions. But they are often disillusioned to note that their recommendations which are fruits of their hard work do not feature in most decisions, as if no evaluation was ever carried out. "A number of writers have been urging evaluators to understand the decision-making systems in organisations and policymaking system in government if we want our evaluations to have any influence. The traditional posture of our field towards policymaking is obsolete. We have been excessively ritualistic and we have presupposed something very much like a benevolent-despot decision maker who would listen to our results and proceed to put them into practice" (idem: 172). She states that evaluation can only change what the public takes for granted when its findings get into currency by clarifying the extent of

problems and by showing alternative routes of action as potential solutions to problems encountered by users. Nevertheless, she makes a concession with regards to the role played by what Patton calls the 'personal factor' (personal interest, commitment, enthusiasm) in determining the degree of influence an evaluation will have, however, the sticking point is that there is no known reason why people become enthusiastic about evaluation – there are no identifiable characteristics that can be shown to be consistently present in such people to know when the personal factor is there, she argues.

In what appears to be a thinly veiled attack on Patton's utilization-focused evaluation, Weiss writes: "You've heard the advice; identify the key decision maker (or the few decision makers). Get them involved in planning the evaluation. Be sure the evaluation addresses the questions they raise. Limit the study to variables that they have the authority to manipulate. Communicate results early. Talk to them in person. When you write, make the report short, clear, and simple. Make explicit recommendations for improvement. Be sure that the results, and the recommendations that you draw from the results, are feasible within the constraints of the organizations" (idem: 172). Despite adherence to this advice, evaluation use has been minimal, she claims. This runs counter to rational decision making process which seems to inform much of Patton's direct planned use. What Weiss is implying is that Patton is falling into the same trap of the positivist tradition that anticipated people to use their rationality in applying evaluation findings because of their purported positive contribution to improvement.

Stakeholder ignorance – Fourthly Weiss argues that even where there are clear decisions to be made and identifiable decision makers, people are often not privy to information they are supposed to know, which is useful in making decisions. They do not have the necessary knowledge to identify information that they need to make a difference. "It is amazing how difficult for them to foresee what kind of information would make a difference. They are inclined to fall back on answers that seem socially acceptable in the organization" (Weiss, 1988a:174).

Fear of change – The fifth and last reason Weiss propounded for the shortfall of traditional advice and evaluation's 'indifferent success' is related to fear of change that evaluation may bring about. Weiss argues that clients, administrators and staff, usually have concrete interests in programmes that they do not want to see anyone change. An

evaluator who comes up with findings and recommendations that imply changes to the *status quo*, is perceived as a threat and the tendency is nonuse of evaluation findings. Weiss claims that this is the reason why major users of evaluation are not those in the programme, that is, use occurring at a distant site from the programme being evaluated.

Process use – In addition to their differences regarding utilisation of evaluation, Weiss and Patton also differ on process use. Weiss and Patton are on the same side on the types of use with the exception of process use. To contextualise it, Patton defines process use as "ways in which being engaged in the processes of evaluation can be useful quite apart from the findings that may emerge from those processes" (Patton, 1997: 88). Advantages of processes of evaluation are incontestable to her; however, she disputes the categorisation of process use as a kind of use. She is persistent that a distinction must be drawn between a type of use and a source of use. Following on Kirkhart (2000), Weiss had the following to say about discussions on process use of evaluation: "This was not so much a new kind of use as it was a new source of use, drawing on something other than the findings. …Process use made explicit an idea that had been implicit for a long time. But it is not parallel to the other three. Instrumental use is presumed to yield decisions of one kind or another. Conceptual use yields ideas and understanding. Political use yields support and justification for action or no action. Process use tells how evaluation's influence arose" (Weiss, 2005: 14). Karen Kirkhart (2000) indicates in her integrated theory of influence that process use was not a type of use. She writes: "Process use first emerged in the utilization literature as a means of facilitating results-based use. Greene's focus (1988b), for instance, was on creating conditions conducive to results-based utilization, rather than on the intrinsic effects of the evaluation process itself. Subsequent treatments of process use emphasized its value independent of results-based use (Whitemore, 1991"" (p. 10). Process use is contrasted with 'results use' and accordingly placed under sources of influence together with findings as process-based and results-based uses.

Patton's response to Weiss: An argument for planned, direct and immediate use

Weiss' presentation awoke the ire of Patton who perceived the presentation as a negative picture of the evaluation profession. He responded by pronouncing that he took "strong exception" to Weiss' declaration and that would be manifested in his presentation the following day in an effort to refute her declarations and delineate the efficacy of evaluation. The following morning, he countered Weiss' with an alternative

vision, a utilisation-focused vision of accountability. This was the beginning of an illuminating debate between the two leaders in evaluation research.

A utilization-focused vision of accountability

In his paper "The Evaluators Responsibility for Utilization" (1988a), Patton presents a case for direct, short-term utilisation of evaluation recommendations. He argues that the necessity for a dissimilar vision was made more pressing, by the 'high esteem' in which Weiss was held by the profession and the perceived negative influence her utterances might have on it. Patton begins with a summary of the essence of Weiss' vocalizations thus: "The overall impression Carol Weiss left on me in her speech was that working carefully and strategically with intended users of evaluation has had "indifferent success", that specific evaluation contributions to programme improvement are rare, and that our best hope is indirect influence "at some distance from the original site" (Patton, 1988a:9-10). A concession is made that Weiss' sentiments may hold in the case of policy evaluation, but with regard to programme evaluation, Patton vehemently refutes "indifferent success" utterance, instead citing evidence that points to the contrary in the utilisation work of fellow colleagues. This evidence, he claims, is not limited to small, local programmes as some evaluation scholars have suggested, but it includes large scale programmes such as the General Accounting Office and the evaluation unit of the FBI, which are national programmes in the United States of America (idem: 10). He paints a picture of divergent visions of policy and programme evaluation which labels Weiss' vision as dismal whilst he shows evidence of success of utilisation and marketability. He writes: "the Weiss alternative of what we can hope for is one of limited vision that I find quite dismal. It is a vision of indirect influence at some distance from the original site, usually at a time much later than the completion of the evaluation, through enlightenment of a fluid, overlapping, ambiguous policy community"(idem: 11). This vision, it is argued, may pertain to policy evaluation, but not to programme evaluation. Patton argues that it is precisely because of Weiss' view of utilisation removed from the programme being evaluated which makes it unmarketable and unaccountable, that it is detrimental to the profession. Clients want something that is potentially useful to their businesses or programmes. He moves to pronounce marketability as a criterion against which utilisation can be judged. It appears as if his strong exception to Weiss' speech is partly connected to what he perceived to be an attack on utilization-focused evaluation, hence its defence that "I can

without difficulty name at least 25, probably 50, practitioners of evaluation who are making good living following utilization-focused evaluation advice"(idem: 11).

The Caribbean Agricultural Extension Project evaluation occupies the pride of place in his heart as an example of a success story of traditional advice on evaluation utilisation on a large scale given its ambition to cover seven Caribbean countries. "The Caribbean Agricultural Extension Project evaluation is one such example, an evaluation of a project aimed at agricultural development in seven countries, an evaluation that directly and importantly affected a $5 million funding decision"(idem:11). Having found fault with Weiss' vision of evaluation, he compares and contrasts his vision with Weiss', effectively acknowledging her vision but adding to or extending it. He writes that:

> "Weiss laid out before us four possible uses of evaluation: warnings, guidance, reconceptualization, and mobilizing support for vested interests. Notably missing from that list is directly influencing change, directly affecting decisions, immediate programme improvement, and follow through of specific recommendations for improvement. My vision includes, indeed emphasizes, such uses" (idem: 12).

The vision is demonstrated through some examples of success of utilisation on a broad scale that come from the very conference (AEA Annual Conference – October, 1987) with the purpose of demonstrating that evaluation utilisation is neither 'indifferent' nor limited to small programmes. One such success story comes from the Government Accounting Office, which reported to have produced 290 reports and 1,134 recommendations. These were followed up through a study on the impact of recommendations. Patton had the following to say about the impact studies:

> "They found that 80% of the recommendations were accepted by the agencies to which they were directed. That is clear, dramatic use of evaluation findings. Indeed, I understand from senior GAO colleagues that their performance evaluations are tied to the utilization of their findings and they are accountable for that use" (Patton, 1988a: 13).

The acceptance rate is no doubt significant; however, it would be interesting to find out how many of those recommendations accepted were *actually* used. Perhaps Patton should have explored a step further to live up to his dictum that 'evaluations should be judged by their utility and actual use'. Nevertheless, the fact that employees/evaluators

are evaluated on the basis of utilisation of their findings is a strong motivation or coercion for accountability for use. In the FBI, evaluators went an extra mile by publishing annual reports of what transpired to their evaluations after follow-ups six months after recommendations were made. According to Richard Sonnichsen, Deputy Assistant Director in the BFI Evaluation Unit at the time, they conducted 37 evaluations of major BFI programmes between 1980 and 1985 with 342 recommendations made. "82% of the recommendations were approved by senior administrators and 90% of those approved were fully implemented" (Patton, 1988a: 13). This is an explicit case of evaluator accountability for use and actual implementation of recommendations. The last example concerns the work by Marv Alkin on the evaluation of educational institutions in California for the State Office. Patton writes that "the State Office of the Extended Opportunity Programs and Services operates in over 100 community colleges and annually publishes a list of all the recommendations made for all community colleges, including what happened to those recommendations"(idem:13). These examples, although not indicative of actual use in so far as recommendations are not followed up or seen to have been implemented usefully (not misused), clearly show a degree of accountability on the part of evaluators by publishing them.

Whether or not these examples answer adequately Weiss' issue of multiple interests in contention, is subject to interpretations. The scale of these evaluations is undoubtedly broad and therefore gives one reason to believe that they included multiple stakeholders with varied interests of the magnitude expressed by Weiss in her argument for 'indifferent' use. Patton would most likely go the route of specific primary intended users in line with the argument that trying to please everybody ends up pleasing no one. Therefore, to achieve high impact, an evaluator has to narrow down stakeholders. He defined utilisation thus: 'By utilization I mean *intended use by intended users*"(Patton, 1988a:14; Patton, 1986). If this definition is considered, then it will not be far-fetched to assume that these examples are meant to be a verification of utilization-focused vision of accountability, where utilisation means intended use by intended users, a direct, planned and immediate instrumental use.

These examples are not only indicative of evaluators' accountability, but also 'the evaluator's responsibility for utilization'. Patton is against the hypocritical stance of those evaluators who demand immunity from accountability/responsibility for their evaluations, but expect it from programme staff. The very activity of being evaluated

has accountability implications, hence he declares: "I'm suggesting that we apply to our professional practice the same demands we make on programs. Moreover, whether we apply those standards or not, it is clear that the funding community and the programme people are going to apply those criteria to us" (idem: 13). Although these examples are testament to instrumental use, one senses an element of coercion in them, particularly with regards to the GAO example where performance evaluation was tied to utilisation of findings. The kind of use that resulted there was in the same vein as what Weiss (2005) refers to as *imposed use*. She talked of imposed use in the following manner:

> "In these stories you will see instrumental, political and conceptual use, and also a new type of use, previously unaccounted for – probably because it is making a debut on the evaluation scene. …it is a type of use that came about because of pressure from outside. In this article, we adopt the term *imposed use*, with an occasional nod to carrot-and-stick use. What distinguishes this use is that program stakeholders are obliged to pay attention to evaluation results"(Weiss, 2005: 16).

The writing was on the wall as the utilisation crisis of the 1970s implied and Patton was merely urging compliance in the light of the profession's dependence for existence on stakeholders mentioned – funders and programme people. Patton sets apart programme evaluators from pretenders through the use of a fishing metaphor – 'a fisherman is the one who catches the desired fish' and likewise, a programme evaluator gets his/her findings used (Patton, 1988a:15). The route to this accomplishment is long and winding. You have to identify intended primary users, establish and foster commitment to use. Involve them in all stages of evaluation while training them, disseminate recommendations and follow them through. Some constraints in conducting an evaluation have to be considered, for example, time for both the expected life-span of an evaluation, that is, when results are expected and the time to be sacrificed by stakeholders. In the first place, an evaluator does not only identify intended users and engender commitment to use, stakeholders must be trainable in the complexities of evaluation and its utilisation. This goes back to Weiss' question about the absence of conspicuous characteristics in stakeholders to identify if they are committed to use, in this instance, trainability. Training also has implications for time and charges of evaluations. When all these factors are taken into account, utilization-focused evaluation is not always a viable option, therefore, there is a trade-off between it and other evaluation approaches depending on the purse, the urgency of results, among other determinants. What is not clear from Weiss is whether any given single policy

formulated by policy makers is an amalgamation of inputs from all these interest groups: mass media, commissions and panels, issue networks and policy staff; or individual policies cover certain interest groups. I am inclined to believe that policies are formulated in such a manner that they concentrate on a certain interest group or related groups, otherwise, they would be vague in focus, thus, becoming less useful and subject to constitutionality wrangles.

Rationality of decision making – Patton's response to this haphazard manner in which certain decisions 'happen' and that decision making is not always a rational process is not direct, but it is covered in his argument for evaluators' responsibility for utilisation, which follows below. While Patton tentatively concedes that Weiss views might pertain to policy evaluation, he likes to take control of the situation from the onset and drive the evaluation process in the direction that will bear fruits. Of course this needs an assortment of skills in addition to methodological training including interpersonal skills, people management, good character judgement, charisma and so forth; but when well executed, it has the effect of rationalising decision making. This is the scenario where decision makers are sensitised to the significance of the process they are about to be engaged in at the beginning of evaluation. Commitment to use is engendered throughout the evaluation process, they are made to own the process by partaking in decision making and fears are allayed through positive expectancy. These can forge a rational connection between evaluation findings and decision making. A critical condition is that funders have to be decision makers and be involved in evaluation.

Stakeholder ignorance – One of the strategies for delivering accountability Patton put forward is *being situationally responsive as an expert evaluator***.** He claims that situational responsiveness goes beyond just attending to physical milieu. It requires that an evaluator also be an expert in human relations in order to be correctly clued to what people expect from them at different points in time judging from both their oral and body languages. It also means being able to educate stakeholders about and assist them conceptualise their needs. Judging from the example Patton used, the expertise required the range of interpersonal skills involved in situational responsiveness (situation recognition and responsiveness, anticipation, and being able to analyse people and knowing where and how to focus attention) is a too expansive for easy mastery. Interpersonal skills required for mastery of situational responsiveness are in fact a

discipline on their own in addition to a host of other skills that are also not necessarily method related. This is the 'no pain, no gain' principle that the Olympic decathlon champion Bob Richards once articulated that Patton seems to have embraced (Patton, 1988a:19). Methodology is supposedly a cornerstone of evaluation, but its significance is seemingly diminished by a quest for utilisation. Case studies may help somewhat in exposing learner evaluators to various experiences of expert evaluators on their path to expertise, but expertise and experience are cumulative as well as practical achievements not to be acquired through formal training only. It takes years of practice for one to be an expert. Having traversed the road strewn with difficulties to accomplish expertise, the evaluator has to educate the ignorant stakeholders as part of what Patton calls the 'evaluator's responsibility for utilisation' in order for them to be able to make correct decisions or choices. Process use is perhaps the greatest tool in this regard. Described as "individual changes in thinking and behaving that occur among those in evaluation as a result of the learning that occurs during the evaluation process" (Patton, 1998:225), process use can go a long way to enlightening those who cannot make the right decisions. It teaches them to think evaluatively. *Learning to think evaluatively* capacitates not only stakeholders but also contributes towards the overall development of an organisation. Advantages of process use have been dealt with extensively in chapter two; however, learning to think evaluatively suffices in this instance of educating people. Of course, process use was not part of the debate in 1988, at least conceptually, because of its late discovery or awareness. Nonetheless, that does not detract its presence and benefits, especially in the light of the participatory nature of utilization-focused evaluation.

Fear of change – Patton (1988a) concurs with Weiss' sentiments on this issue from experiences with staff who were either hostile towards or fearful of evaluation. However, unlike Weiss, he is not content to acknowledge that only he offers solutions to allay reasonable fear among those accountable. One of the strategies for a vision for accountability is 'overcoming staff fear'. He goes about it thus:

> "One of the things I to do is deal with the fear of evaluation up front. I assemble the primary users, administrative representatives, funding representatives, staff representatives, and client representatives for an initial session where we talk about what the evaluation process is going to mean, and begin to engender commitment to use"(1988a: 16).

It is in such sessions where stakeholders raise their concerns about and experiences with evaluation and evaluators. The evaluator also gets to gain insights into their mode of reasoning, the programme culture and other contextual issues that have to be considered for a successful evaluation. S/he contributes to the session by projecting a positive image of evaluation that allays fears of stakeholders and creates positive anticipation among them. They therefore look forward to a useful evaluation. This is apparently not for an average evaluator and it involves expertise in change management, conflict resolution as well as creativity.

Academic versus service evaluators

Patton followed that presentation with an article in his President's corner of the May 1988 issue of *Evaluation Practice*, *"*How Primary is your Identity as an Evaluator?". In the article, Patton rationalises the divergent nature of his views and those of Weiss in terms of the results of a study conducted by Shadish and Epstein (1987) on the internal dynamics of evaluation profession of members of Evaluation Network and Evaluation Research Society which had since merged to form the American Evaluation Association. The significance of that article to Patton is that it came immediately after the Boston conference and 'it helped me understand some of what transpired there' (Patton, 1988b: 205). Of particular interest to him were the responses to a question on professional identity. Shadish and Epstein write: "However, we also asked if respondents would personally describe their professional identity as evaluator, and 31% of the sample said yes" (Shadish & Epstein, 1987: 560). The study also indicated that of over 200 members who answered all questions, only 8% were trained specifically as evaluators, with more than 60% having received training in psychology and education. The result depicted evaluation as largely an add-on activity to most evaluators' primary professions. It might not even had been practised independently, but as part of the routine performance of their primary functions. For instance, the principal's evaluation of staff and student performance is their duty as management of their profession, education. This is indicative of the evolution of evaluation from other professional fields. Owing to the influence of context or one's angle of vision in relation to the object of study, the findings of this study on professional identity of evaluators assisted Patton in understanding the differences in his and Weiss' view of evaluation. He argues that their differences can be attributed to Weiss' academic posture against his practical disposition. He declares that:

> "They found that responses clustered around two contrasting views of evaluation: "academic versus service evaluators"(p.587)…. "Academic evaluators" tend to emphasize the research purposes of evaluation, traditional standards of methodological rigor, summative outcome studies, and contributions to social science theory. "Service evaluators" tend to emphasize serving stakeholders' needs, program improvement, qualitative methods, and assisting with program decisions" (Patton, 1988b: 204).

The significance of the differentiation for Patton was that it enabled him to rationalise their differences between himself and Weiss in terms of academic and service evaluators respectively.

Shadish and Epstein went further to outline the potential danger of these differences to evaluation practice. The warning is grounded in the schism that occurred in psychology between academics and practising psychologists due to similar orientations found in evaluation. They aver that "this diversity also brings with it potential increased tensions between applied and academic interests, tensions that arise because of the different demands and reward structures under which the two groups often operate"(Shadish & Epstein, 1987: 587). Patton seems to heed the warning when he 'puts on record' his respect and advocacy for membership diversity in the American Evaluation Association. He sounds apologetic about his over-enthusiasm in dismissing Weiss' perspectives because of their divergence with his 'service-oriented' stance when he says: "I may have appeared intolerant of those who do not share that view or engage in that practice. This was not my intent. My intent was to facilitate open discussion and debate about alternative visions of evaluation practice" (Patton,1988b: 206).

Perhaps he wrote this piece with the benefit of hindsight because it is not far-fetched to claim that his dramatic and animated response to Weiss' presentation bordered on narrow-mindedness or was intolerant. He is in fact economical with the truth as the tone of his response on the AEA conference attests. He wrote that Weiss' perspective about evaluation was "a vision that I personally found to be quite dismal, sad, and unacceptable" (Patton, 1988a:70). I find it difficult to extract anything that remotely promotes debate or accommodative in this utterance, it is completely intolerant.

Weiss is not at ease with Patton's insinuation that as an academic, she cares less about evaluation use than practitioners. She responds: "Drawing on a study by Shadish and

Epstein (1987), Mike (Patton) suggests that there two kinds of evaluators: those who are interested in working with clients to maximise use of their results and those who are "academic" in orientation and care little about use. He implies that I am in the latter variety" (Weiss, 1988b: 220). She counters that while it is a fact that she teaches at a university, she is 'as committed to the use of evaluation results as any person alive. On the issue of dichotomies - judgement versus service, truth versus action, outcomes versus processes, she responds thus: 'I am interested in service and don't eschew judgement; I believe truth helps to foster action'. She says her experience has taught her that people are more inclined to use research if they are convinced that it is accurate and defensible. She also expresses hope that the time has come to outgrow the habit of categorising human behaviour into dichotomies (good and bad, academic versus caring *etc*.).

Weiss' response: An argument for influence

Weiss followed up her earlier contribution to the debate with another one in response to Patton's animated presentation in a paper she gave at the same conference and to his follow-up in which he branded her as an "academic" evaluator more interested in theory-generation than in evaluation practice in "If program decisions hinged only on information: *A response to Patton*" (1988b). Weiss points out what she perceives to be the strengths and weaknesses in Patton's arguments and she highlights the implications of their positions.

Indifferent success – On this point, Weiss maintains her position that evaluation results had 'indifferent success' in making evaluation the basis of decisions. She contends that despite the cases Patton and other evaluators may name, they do not refute her point that 'people do not routinely use evaluations as the bases of decisions'. For her, even those who use utilization-focused evaluation, their influence is good at 'about midpoint on the rating scale' (Weiss, 1988b: 210). While she encourages evaluators to do more to draw attention to evaluation findings, she is sceptical that evaluators can persuade stakeholders to make evaluation results the 'overriding consideration in programme decisions'. This perspective stems from her conviction that programme people know more than what the evaluator tells them. She declares:

> "The evaluation cannot cover all aspects of the program, and it can never be the basis on which decisions are made. To ask program managers and planners to embrace evaluation findings

> fully is to ask them to bracket their years of experience and direct immersion in the daily world of the program, in effect to abdicate their responsibility in favour of the evaluator who inevitably has only a partial view of their dilemmas" (idem: 211).

Evaluation, she argues, should be geared towards adding better understanding of programmes, giving options and likely effects. Therefore, it should aspire to *influence* instead of utilisation. Obstacles to that are outlined hereunder. She does not attempt to repudiate Patton's claim, at least at the beginning, that he could *'without difficulty'* name at least 25, probably 50 practitioners of evaluation who are having dramatic impact on programmes and policies but questions that number Patton mentioned in relation to the number of evaluations conducted up to that point in time. She clarifies her standpoint that she did not say that use was bad through a reference to Webster's Collegiate dictionary where 'indifferent' means "neither good nor bad, large nor small, desirable nor undesirable, etc"(idem: 210). In the light of this explanation and the number of evaluations Patton spoke about (25-50), Weiss asks: 'Out of how many? 100? 200?' Unfortunately, Weiss gives only a tentative estimation of the number of evaluations carried out against which success rate of Patton's 25-50 (also tentative) can be determined. Reliability of their estimations is made more suspect because of their hundred percent jumps, from 25 to 50 and 100 to 200 respectively, estimated numbers are supposed to be closer together not poles apart. Nonetheless, if one takes Patton's highest estimate (50) against Weiss' lowest estimate (100), utilisation impact success would still be 'indifferent' if Webster's collegiate dictionary is followed. If one takes the highest estimates from both luminaries (50 and 200), then utilisation of evaluation was worse than 'indifferent'. Therefore, one is inclined to agree with Weiss that she intended 'indifferent success' to be a beacon of hope. She writes:

> "What I wanted to convey – and still do – is that even when evaluators try to follow utilization-focused prescriptions, then influence is okay, fair, and about midpoint on the rating scale. I think evaluators can do better than they are doing to encourage attention to their findings, and I think they should. But overall, I doubt that we can ever persuade stakeholders to make evaluation results the overriding consideration in program decisions. For one thing, program people know more about their programs than simply the things the evaluator tells them" (idem: 210).

A closer look into Weiss' assertion reveals that the bone of contention is not utilisation per se, but the extent of utilisation and the manner in which it happens. Weiss is at odds

with Patton's compassionate affiliation to direct almost linear use as manifested in intended use by intended users just as the latter is against the winding road to use that the former traverses. The language used by Weiss is worth noting because it is a reflection of the gulf between her and Patton. She speaks of *influence* to connote distance from the programme being evaluated while Patton speaks of use. She writes passionately that:

> "What evaluators should aspire to achieve in the area of utilization is influence, not the status of philosopher-kings whose dictates determine program future. It is presumptuous to think that one evaluation study, no matter how conscientiously done, should be the major basis for changes in program. (I won't even mention the evaluation studies that are done at mediocre levels of competence or less.) Evaluation is better advised to add to understanding about the program, to illuminate the range of options and likely effects. In essence, evaluation should be continuing education for program managers, planners, and policymakers" (Weiss, 1988b: 211).

She proceeds to outline some of the reasons why Patton's utilisation-focused advice and a multitude of interpersonal skills do not suffice for an 'evaluator to have dramatic impact on the decisions of the rational and high-minded managers and policymakers'.

The ubiquitousness of programme politics – Impediments to evaluation use are further delineated that even the less loaded *influence* becomes difficult to be realised. She claims that people in programmes have many interests that may prevent them from paying close attention to evaluation results. "They want to have satisfactory work life, get along with their colleagues, gain recognition and respect, see their organization grow in prestige and financial solvency, maybe outdo a competing agency or faction, have a chance for advancement, do work that is esteemed by members of their profession, observe valued rituals, and have fun" (idem: 211). She pronounces that when evaluation results put some of these values in jeopardy, programme managers and staff sometimes give precedence to their own interests. She points to the absence of reference to politics in Patton's response to her presentation and other related issues such as availability of resources, avoidance of unpleasant tasks, programme reputations, getting larger appropriations *etc.* She then gives examples of how these affect evaluations.

- At one stage she was a director of a programme financed by federal government in a social action agency. The evaluation a report was made and the programme

was quite successful, but no local agency came forward to pick up the programme. The success of the programme is reported to have instilled fear in the staff of the sponsoring unit about their future to an extent that financial support for the programme was withdrawn. The result was that successful findings were not utilised.

- Another case involves an evaluation study where staff was divided about allocating generous resources to one programme at the expense of others. An evaluation was carried out and findings indicated both advantages and disadvantages of such a decision. Divergence of opinion persisted as strongly as before, albeit with support from evaluation findings. Each side cited evaluation evidence which pointed the strength of their argument and the weakness of the other to justify its deportment.
- There was a million-dollar evaluation in which she was involved as a consultant which she claims provided useful information about both processes and outcomes. It also made recommendations about improving programme effectiveness. Unfortunately for the programme, a new person was appointed as head at the parent agency when the programme was about to finish. His priorities did not include the programme mentioned, the director was forced to leave and consequently, the programme was sidelined.
- One evaluation found one mode of programming to be better than others, and it recommended its expansion. The problem was that it was less staff-intensive and its costs were 30% more than usual programming. It would also serve less people than usual programming. The negative publicity associated with the rising costs and cutback in the number of people (as opposed to cost-effectiveness and job creation) made the agency to proceed as before, only making minor adjustments.

Understanding how organisations work – Weiss claims that agency staffs have an inclination to act non-rationally. The rational process of 'defining problems, searching for alternative solutions, and choosing the most satisfactory alternative on the basis of evidence' is seldom seen by scholars of organisations, she contends. "Rather, they see organizations acting according to bureaucratic rules and standard operating procedures, or through organizational politics (with factions or subunits vying for advantage), or through "garbage-can" processes (where a decision is the almost-chance confluence of the streams of participants, problems, solutions and chance opportunities that are

flowing through the organization at the moment) or acting first and crafting explanations of their actions afterward, retrospectively labelling them "decisions"" (idem: 214). She asserts that there is consensus among organisational scholars that organisational members' welfare takes precedence over the people they serve. She then turns attention to some of the issues raised by Patton.

Reliability of evidence – The reliability of evidence that Patton uses to determine the level of utilisation of evaluation comes under attack. Weiss asks: "On what does Mike base his belief that the level of use of evaluations is high? Mostly on the basis of evaluators' accounts of their own success" (idem: 216). She writes that unlike a practitioner whose job is to believe, the job of an evaluator is to doubt, to examine the evidence. The evidence presented by evaluators should also be subjected to careful scrutiny to see the other side of the story, not just the successful side, she persists. Another type of evidence on which Patton relies is the 'follow-ups' by evaluators on their recommendations. Although she says she respects them, she is doubtful when it comes to taking recommendations as units of utilisation. "To the best of my knowledge, we haven't looked very closely at where evaluators' recommendations come from. Some recommendations may be well grounded in data, and others may be flights of fancy from people without much expertise in program planning or operation" (idem: 217). This suggests that disjunctures may occur between data and recommendations. Sometimes when programmes are not doing well, recommendations are made logically by evaluators based on what they think or know may work better, in the process closing gaps and fixing inconsistencies, she argues. But I sense a problem whereby any presentation will be regarded with dubiousness unless the evaluation it is based on has been subjected to meta-evaluation. This has implications for the findings presented to those who commissioned evaluations, whether or not they should also wait for meta-evaluation because if evaluators' account of their own success is in doubt, recommendations cannot escape it.

On competing paradigms – Weiss is uneasy with Patton's insinuation that as an academic, she cares less about evaluation use than practitioners. She responds: "Drawing on a study by Shadish and Epstein (1987), Mike suggests that there two kinds of evaluators: those who are interested in working with clients to maximise use of their results and those who are "academic" in orientation and care little about use. He implies that I am in the latter variety" (Weiss, 1988b: 220). She counters that while it is a fact

that she teaches at a university, she is 'as committed to the use of evaluation results as any person alive. (As a precondition, I am also committed to improving the methodological quality of evaluations, both quantitatively and qualitatively, so that they have something valid to say.)' (idem: 220)..

A vision for the future – a few things that she hopes that they can agree on with Patton are stated: For instance, improving the quality of evaluations, asking the right questions, improving methodological, people and programme life skills, improving analysis and interpretation skills and encouraging evaluative cast of mind among policy and programme planners. But she cautions that they should not expect skills and goodwill to end self-interest, political advantage and other obstacles to use or evaluation influence. She ends with an excerpt from Rossi (1987): "In the short term, good social research will often be greeted as a betrayal of one or another side to a particular controversy [and rejected]. …It is important to keep in mind that in the long term, good research will often, if not always, be assimilated into conventional wisdom" (Rossi, [1987] as cited in Weiss, 1988: 221). This is in some way a form of validation of her standpoint because the extract is effectively an argument for influence in the long-term.

Rationalisation of the Patton-Weiss through Habermas' purposive-rational and communicative actions

Smith and Chirop (1989), employed communicative action theory as expounded by Jürgen Habermas to rationalise the seemingly irreconcilable stances that Weiss and Patton took during their debate on evaluation utilisation. They commence by noting that both Patton and Weiss through two articles each (Patton, 1988a & b and Weiss, 1988a & b) presented 'competing views of both ways in which evaluation can influence decision makers and the actual level of impact of evaluation on decision making'(Smith and Chirop, 1989: 5). On the one hand, Weiss argued that following traditional advice for conducting useful evaluations had not noticeably increased evaluation impact on decision making. She writes: "Even evaluators who have tried conscientiously to abide by the traditional advice had indifferent success in making evaluation the basis of decisions" Weiss, 1988a: 7). On the other hand, Patton was strongly upset by Weiss' pronouncement that utilization-focused evaluation advice results in 'indifferent success' and the assertion that even when traditional advice influences decisions, it is 'often indirectly and at some distance from the original evaluation site'. In response, he

categorised Weiss under 'academic evaluators' who are more interested in those aspects of evaluation that will contribute towards knowledge accumulation for theory building or in passing judgement. He leans on the work of Shadish and Epstein to illustrate his point that " "academic evaluators" tend to emphasize the research purposes of evaluation, traditional standards of methodological rigor, summative outcome studies, and contributions to social science theory. "Service evaluators" tend to emphasize serving stakeholders' needs, program improvement, qualitative methods, and assisting with program decisions. …Carol Weiss has an academic research perspective on evaluation. I take an action research and stakeholder-oriented approach" (Patton, 1988b:88-89). However, Weiss is adamant that "I am as committed to the use of evaluation results as any person alive"(Weiss, 1988b: 26). She counters that Patton is apolitical because of his failure to mention 'politics' in his manuscript. She declares "self-interest, organizational protection, the quest for advantage – these elements are noticeably absent from the Patton world. In a manuscript about the use of evaluations that is 25 typed pages long, Mike Patton never mentions the word "politics". In his world everybody behaves rationally. …They don't seem concerned about drains on the budget, finding qualified staff, extra work, disturbing ongoing relationships with other agencies, possible negative feedback from community groups or the press, getting the grant renewed, satisfying the curmudgeons on the board, or any of the concerns that exercise the program people I have dealt with" (Smith and Chirop, 1989: 7-8).

Smith and Chirop believe that 'both authors are right', and they elucidate areas of common agreement between them before expanding on their points of divergence. Their common ground includes their concern for utilisation of evaluation, to increase it and they acknowledge the multiplicity of stakeholders in decision making processes. A disturbing feature, however, is the authors' disregard for stakeholders on whose behalf, programmes and policies are made, they note. "Further, neither author seems much concerned with the "end-of-the-line" stakeholder, the customer, the public. These evaluators are preoccupied with decision makers: managers and policy makers, the "movers and shakers" of society…" idem: 7). It is apparent that accusations and counter-accusations between the two luminaries of evaluation that one is 'academic' and the other is 'apolitical' do not stick. A question then arises as to the reasons for such divergence of opinion whilst they share the same fervour for utilisation of evaluation results. The authors attribute different postures taken by Weiss and Patton on utilisation to different decision-making communities they serve and their rationales.

They write: “Weiss (1988a) characterizes her decision-making context as one in which decisions result from contention among solidified interests; where decisions often do not result from rational decision processes, but seem to “accrete”, “where authority is dispersed, multiple groups have a say, and policy is the resultant of conflicts and accommodations across a complex and shifting set of players”” (Smith & Chirop, 1989: 7). They argue that in Weiss’ world of policy making, an evaluation is only one of the many interests contending for recognition, all with justifications from their proponents. In this world, they write, there is no managerial control, no agreement on means and on ends and decision making is not rational.

They note that “Patton describes a primarily cooperative environment of a more unified community of shared values and common goals. More direct managerial control is possible, and with greater agreement about desired ends, more effort is focused on finding suitable means – an environment in which technical rationality is more effective” (Smith and Chirop, 1989: 8). They claim that technical rationality is instrumental to the development of and distinction between alternative means to agreed-upon ends; it is more effective in Patton’s world because of the widely shared values and the converse is true in Weiss’ because of widespread disagreements. Furthermore, technical rationality is said to work well with issues concerning means and ineffective with regard to ends, therefore, it works better where there is common ground on ends. In Patton’s world, there is relatively less contestation of values compared to Weiss’ world, which makes ends easier to decide – consequently, technical rational is more effective. They then elaborate on the connection between reason and action using Habermas’ claim that ‘knowledge and human activity are constituted and driven by three interests: purposive-rational action, communicative action and emancipatory action’ (idem: 9). Purposive-rational action is goal-oriented and it is subdivided into two. One – instrumental action, which follows rules based on empirical knowledge (this determines appropriateness and is means-ends rationality) and two – rational choice based on analytical knowledge (here inferences are based on preferences and decision rules, though partly social, it is still based on means-ends rationality). Communicative action is ‘person-oriented social action. It is governed by concensual forms that define reciprocal expectations and is grounded in intersubjectivity’ (idem: 10). The authors put forward that purposive-rational action (related to technical competency) is enough for evaluation practice in Patton’s world, but both purposive-rational and communicative actions (related to societal norms) are

required in Weiss' world. They further examine the kinds of truth claims that the two kinds of action make.

> "A truth claim under purposive-rational action is supported by an empirically true or an analytically correct proposition. Punishment is built into the system. Failure to observe the proposition will lead to failure. Here failure is similar to incompetence or lack of efficiency…. Truth claims under communicative action are supported by consensually achieved social norms (intersubjectivity). Failure to achieve these claims is sanctioned by the conventions of the community. The punishment for being "deviant" is external to the proposition and grounded in society" (idem: 10).

Patton's insistence on accountability on the part of evaluators for non-use of evaluation is regarded as testimony of his affiliation to purposive-rational action because failure to do so, is incompetence. Weiss's readiness to attribute utilisation responsibility to both evaluators and to a host of other players (society) introduces communicative action in addition to purposive-rational action. Therefore, Weiss' world is broader than Patton's.

The authors argue that restricting evaluation to rational-purposive action, focused on technical application, 'is not very different from an exchange of money in an economic system where barter of goods is the only determinant' (p. 10). For them, evaluation should influence decisions about ends and means and purposive-rationalist framework is silent on the choice of ends – it only addresses how to reach them, not how they are chosen. Moreover, they argue:

> "Questions of ends involve goals and values and are beyond purposive-rationalist framework; the communicative action framework must be invoked. At this level, interaction takes place among human subjects and thus consensus is needed. Evaluation within political decision-making communities such as described by Weiss, where both ends and means need to be resolved, will require both purposive-rational action and communicative action – the position Weiss seems to be advocating, at least implicitly" (idem: 11).

They align themselves with Habermas' pronouncement that in its quest to emancipate knowledge (praxis) from the rule of the irrational (dogma), the rational-purposive philosophy created an enslavement of a different type. "The expert-client system negates the full emancipation that the rational-purposive approach set out to give.

Knowledge is enslaved both by limiting itself to only those phenomena that can be grasped by rational measurement and by limiting itself to the rule of a few experts who make a monopoly of the situation" (idem: 11). Emphasis on evaluator accountability for failure of evaluation by Patton indicates the centrality of an expert/evaluator in the purposive-rational evaluation in his framework. This framework, though essential, particularly with regard to means, it does not help with problems of ends which require communicative action. The latter framework is not without problems as reflected in the authors' pronouncement: "But although the approach advocated by Weiss frees us from the narrowness of purposive rationalist expert, it introduces its own form of enslavement by expert in its narrow conception of who participates in its decision-making community" (idem: 11). They conclude that "Patton's approach provides us with an essential mechanism for achieving desired ends, once they are determined. Weiss approach supplements that contribution by including not only expert development of means, but also expert selection of ends. Neither of these approaches addresses the problem that control of both means and ends are increasingly concentrated in decision-making communities of experts" (idem: 12). The solution to this is to use stakeholder involvement both to increase evaluation utilisation (purposive-rationalist action) and to counteract the concentration of control of decision making in the expert through open, productive engagement that empowers other stakeholders and that gives them confidence to neutralise the evaluator, thus, decentralising power.

In a response, Patton discounted the arguments raised in this article about their differences being a consequence of their differences in spheres of operation even before it was written, for in "How Primary is Your identity as an Evaluator" (1988), he wrote that:

> "Many people have since suggested to me that Weiss and I are simply operating in different arenas, that her remarks were addressed toward summative policy evaluation at the national level, while mine were aimed at a more narrowly defined version of formative program evaluation. While it does appear that we typically operate in different arenas and are involved in evaluation in different ways, I believed then, as I do now, that we have fundamentally different views of evaluation as a field of professional practice. Carol Weiss has an academic-research perspective on evaluation. I take an action and stakeholder-oriented approach – what I've called "utilization-focused evaluation"" (Patton, 1988b:205).

Perhaps the differences between Patton and Weiss can be located in their modus operandi. Patton, on the one hand, operates in a selective manner by choosing studies he would like to conduct. That affords him opportunities to avoid contentious situations while giving him leeway to apply the utilization-focused evaluation precepts that results in increased use. He writes:

> Operating selectively, as I acknowledged earlier, I choose to work with clients who are hungry for information to improve programmes. They are people of competence and integrity who are able to use and balance both positive and negative information to make informed decisions. I take it as part of my responsibility to work with them in ways that they can hear the results, both positive *and* negative and use them for intended purposes. I don't find them resistant. I find them quite eager to get quality information that they can use to develop programmes to which they have dedicated their energies. (Patton, 2008: 550).

Weiss, on the other hand, is less selective as the challenges she confronts indicate her involvement in some difficult situations. It is not all evaluators who have the elitist comfort of picking and choosing what they like to evaluate as Patton appears to have. If all evaluators can be given the luxury to choose where they want to work, the situation in the profession could deteriorate to a point where only those organisations that are extremely rich will afford evaluations and only a few elite evaluators conduct evaluations. This can lead to the extinction of the profession and societies will be poorer for that. The irony is that Patton sounds like he has the interest of evaluation research and practice at heart yet he relegates the difficult work to others. Henry may be correct in emphasising social betterment as the ultimate goal of evaluation because use usually ends with programme funders or decision makers, administrators and staff who implement evaluation processes and findings. When they are conducted with the suffering of those programmes are supposed to develop, then, it may be morally difficult for evaluators to be selective and intended users will shift to the end of the logic model, improving the human condition. That can force evaluators of Patton's ilk to follow that goal.

6.3 Patton's debate with Scriven on the formative-summative distinction

This debate is based on Michael Scriven's definition of formative and summative evaluations. Patton, 1996 contests the efficacy of the formative-summative dichotomy in capturing the purposes of evaluation as propounded by Scriven in his 1991 paper,

"Beyond Formative and Summative Evaluation" which delineated fallacies about formative and summative evaluation. His contention is based upon the observation that evaluation has grown beyond those two purposes. Scriven's description of the two purposes makes them exhaustive. Scriven wrote:

> "Formative evaluation is designed, done, and intended to support the process of improvement, and normally commissioned or done by, and delivered to, someone who can make improvements. Summative evaluation is the rest of evaluation: in terms of intentions, it is evaluation done for, or by, any observers or decision makers (by contrast with developers) who need evaluative conclusions for any reasons besides development" (Scriven, 1991: 20).

By pronouncing summative evaluation as the rest of what is not covered by formative, Scriven made the two purposes all-encompassing of evaluation's purposes. Patton then outlines three areas of evaluation use where formative-summative distinction is inadequate, namely; knowledge generation, developmental evaluation and the use of evaluation processes to support interventions or empower participants.

Knowledge generation - In contrast to instrumental use, where action is taken as a result of evaluation, knowledge is generated through conceptual use, that is, the use of evaluation processes and findings to influence thinking generally. While formative and summative evaluations do contribute to conceptual use, Patton emphasises the importance of knowledge generation as an independent purpose of evaluation. Among examples cited as knowledge generating/enlightening is *theory-driven evaluation* for its informative insights into general characteristics of successful programmes. According to Chen, theory-driven evaluations can provide models for summative judgements or for improvement. The connection with social science theory also offers the potential for increasing knowledge about how effective programmes work, thus laying the basis for understanding of what produces good results. This knowledge is not limited to particular programmes but is generally instrumental as a guidance for future considerations for programme development and implementation. The same can be said of synthesis evaluations, which synthesise information from different studies to come up with a general pattern for a certain given variable under study. These syntheses represent an accumulated wisdom on particular aspects of evaluation that can be adapted for future use. For example, James Burry's synthesis on factors affecting utilization is a product of an extensive study of those aspects across multiple

evaluations and their utility. Conclusions drawn from such syntheses can be used to inform other evaluation studies. Patton highlights another form of synthesis evaluation referred to as 'cluster evaluations' which were conducted by the Council on Foundations. Cluster evaluation is when an "evaluation team visits a number of grantee projects with a similar focus (e.g., grassroots leadership development) and draws on individual grant evaluations to identify patterns across and lessons from the whole cluster. …The McKnight Foundation commissioned a cluster evaluation of 34 separate grants aimed at aiding families in poverty. One lesson learned was that "effective programs have developed processes and strategies for learning about the strengths as well as the needs of families in poverty""(Patton,1996:133;Patton, 1993:10).

Developmental evaluation - The formative-summative evaluation is also inadequate in covering developmental evaluation, where there is ongoing work for improvement in ongoing programs without a need for summative evaluation. The improvement process is also not formative because the aim is not to prepare the programme for summative evaluation. Patton writes:

> "I have been working for over ten years with some philanthropic foundation programs in areas where they have made long-term commitments and do not want or need to make judgement about continuation. They do want to continue improving the effort over the long-term, however. A variety of evaluation issues have been addressed over these years and each effort has contributed to changes in the program. Moreover, it is the philosophy of the program staff that the program must continue to develop in response to changed conditions and new knowledge. Such changes do not mean that what was done before was ineffective; rather, it means that the world changes, the program must change. Developmental evaluation supports such ongoing development. This is very different from conducting formative evaluation in an effort to get ready for summative evaluation" (Patton, 1996:134-5).

Developmental evaluation shares the improvement orientation of formative evaluation but, it is not connected to summative evaluation. In developmental evaluation, improvement is an adaptation to changing program dynamics or the changing world around programmes or organisations in which programmes are, as well as adjustment to innovations. No outcome is expected from processes, processes are the outcome. This is in contrast to Scriven's original distinction which defined formative evaluation as preparing the programme ready for summative evaluation (idem:135).

Using the evaluation processes as interventions - The other area where formative-summative distinction is insufficient is 'in describing and thinking about uses of evaluation processes', Patton contends. He writes that:

> "As I suggested earlier, the formative versus summative distinction actually concentrates on how *findings* are used. One of the things we have learnt in recent years is that much of the impact of evaluation comes from the process of engaging in the evaluation, not from the findings. Working with a program to understand the logic of evaluation, to develop the program's model, and otherwise engaging staff, through education, to think about and become more effective in their program efforts, can be a purpose of evaluation"(idem:136-7).

Patton provides some examples of how evaluation principles in the processes of evaluation can be useful. One such usefulness is in negotiations among groups whereby the previously opposed parties ended up agreeing on the design of the project and how it should be evaluated. In addition to providing mutual understanding, evaluation can also promote communication by allowing stakeholders from different backgrounds opportunities to tell their stories. This allows them glimpses into what informs one another's reasoning, thus forging tolerance. Being heard can be an end it itself, especially among the those formerly voiceless or disadvantaged – it can mean progress. In participatory evaluations, evaluation processes can take on the function of interventions to increase programme effectiveness, Patton claims. He further argues that when that is the case, "traditional separation between evaluation and programming falls away and the evaluation is conducted in such a way as to support and reinforce the program intervention. This kind of evaluation design has the purpose of directly enhancing the program's impact by the process of engagement rather than the use of findings. I submit that such conceptualization deserves classification beyond the formative-summative distinction" (idem: 137-8).

Complementarity of needs assessment and assets analysis

Patton (1987) is concerned about the amount of attention to needs of stakeholders as expressed in the eminence given to needs assessment, particularly by Scriven, to the almost exclusion of assets analysis. He states that "a number of evaluators, most notably Michael Scriven, have made needs assessment the first and most fundamental step in program development and evaluation" (Patton, 1987:69). Needs assessment

gives attention to the needs of clients in order to meet them, and many programmes are funded based on results of needs assessments. While this is laudable, it is carried out mostly to the exclusion of assets analysis which accentuates the strengths of clients and builds on them. Patton is not averse to needs assessment but he communicates the necessity for assets analysis to provide counterbalance. He writes:

> "Assets analysis is not the opposite of needs assessment. Rather, it highlights a different aspect of the assessment process. Where needs assessment suggests deficiencies to be corrected, assets analysis calls attention to strengths that can be developed. Just as programs are often unaware of the needs of their clients, so too they are often unaware of the capabilities of their clients. Those strengths or assets, if known, can be used to help clients meet their own needs" (idem: 70).

Whereas needs analysis focuses on things that are lacking in programmes, assets analysis gives attention to things that are working well and keeps them. Patton cautions that accentuation of needs and disregard for strengths risks the adoption of a 'deficiency approach' in educational and social action programming. This perspective is pervasive in American educational system and is manifested in teachers' instructional objectives where emphasis is laid on things which students cannot do with a view to impart them – the same approach is used in many programmes, he declares. He explicates the existence of a deficiency approach in the following terms:

> "The deficiency perspective exists when the student or client is viewed largely in terms of deficits to be removed or corrected, as an empty vessel to be filled, or yet again as clay to be molded. Program staff can become so concerned about what is to be done to and for the client that they lose sight of what can be done *by* the client. Clients in such programs come to view themselves in the same way; they become highly aware of their weaknesses and scarcely conscious at all of their strengths" (Patton, 1987: 70).

Mental damage is highly probable when students or clients begin to see themselves in the same terms as exponents of needs assessment see them, students and clients may end up intellectually dependent, unable to think on their own. Furthermore, programme staff is unlikely to meet the needs of their programmes if they do not know their strengths or assets. While assistance is welcome, people should be helped to help themselves and that can be achieved if their assets are identified. It is better to teach people to catch fish than to give them fish because the latter creates dependency and a

sense of disempowerment, to use empowerment evaluation language. Lastly, Patton underlines the fact that statements of assets should be put in such a manner that they do not merely indicate 'non-needs'. He writes: "Deficiency thinking can become so predominant that even strengths are stated as absence of needs. He cites an example of an evaluation report that concluded: "The program needs no new staff because it doesn't have a turnover problem and staff morale is not low." Such a conclusion hardly points out to program staff and others the basic strengths of the program: "The program has a stable and highly enthusiastic staff"" (Patton, 1987: 71). In short, Patton calls for a balance in reporting, pointing out areas of weakness to be improved and highlighting strong points to be maintained.

Scriven's response to Patton

On formative-summative distinction – Scriven's response to Patton's pronouncements on the inadequacy of the formative-summative dichotomy to cover evaluation's purposes, which appears in "Types of Evaluation and Types of Evaluator" (1996), was to clarify and modify his position on the distinction. But he was not impressed with Patton's submissions on other purposes of evaluation. He remained steadfast that knowledge generation, developmental evaluation and process use as intervention fall outside the ambit of evaluation purposes. Scriven commences by explicating that formative and summative are elements of one classification of roles of evaluation, just as process and outcome, research and practical roles belong to other classifications. He then elucidates what he terms 'misconceptions and misnomers' about the formative-summative elements. There is a tendency among evaluators, Patton included (Patton, 1996:132) to speak or write about formative and summative evaluation as occupying opposite ends of a continuum, with formative being associated with improvement while summative is linked to judgement. According to Scriven, the difference is context-dependent and changes with it as in the example below:

> "The guest's reactions are summative in the usual context of e valuating one meal, but can be made formative by changing the context to that of an ongoing attempt by the host or restaurant management to use feedback as a basis for improving the food. …A book reviewer is typically doing summative evaluation of the first edition: but if the book goes to a second edition, those reviews will a good evaluative basis for making changes, hence de facto formative evaluation for the second edition" (Scriven, 1996: 153).

The illumination also lays to rest the misconception that formative evaluation is employed by evaluation practitioners and summative by evaluation by social researchers because they are not as distinct as it is often stated. Scriven declares that when he introduced the distinction, he emphasised that the difference was not intrinsic but contextual, and therefore their roles are defined by context.

With regards to Patton's concern about perceived supremacy of summative evaluation over formative evaluation as expressed in Scriven, 1991 that "the formative evaluation should *at least* provide a preview of a summative evaluation" (Idem: 28); there appears to be a problem of interpretation. This is interpreted by Patton to imply subservience of formative evaluation as in a curtain-raiser before the main match. This is more a problem of interpretation than of fact.

It seems as though Patton had a preconceived idea and wanted something to fit into it because the interpretation does not follow from what Scriven wrote. Patton actually cut Scriven's sentence in the middle, took what supports his argument and left the rest of the sentence hanging. The full sentence reads thus: "Put in another way, the formative evaluation should *at least* provide a preview of a summative evaluation, since one of its most useful functions is to be an "early warning system"" (idem: 28). The subordinate clause is crucial because it illuminates the reason for Scriven's pronouncement which was based on formative evaluation's function and not on Scriven's perception of summative supremacy; it also highlighted another use of formative evaluation, a signal when things go wrong. Nonetheless, Scriven responded that he does not prefer either role, and the issue of supremacy is irrelevant to him. He asserts: "I come to the defense of summative more often simply because more people in the evaluation literature have attacked it or dismissed it or underestimated it; but I do – and enjoy doing – formative evaluation as much or more"(Scriven, 1996: 154). Interestingly, after the above pronouncement, Scriven goes on to defend summative evaluation by citing big events such as the Olympics, Tour de France and Cup Finals where it is relevant and nothing to show for his 'much more' enjoyment of formative evaluation. Perhaps attacks on summative evaluation have been too much to leave them be or a defence has turned into a habit or it may be unconscious mind exhibiting his inclination. Either way, that does not foster trust in his pronouncements about the comparison/contrast between formative and summative evaluation being a 'non-issue'. Paradoxically, Patton agrees with Scriven in the latest edition of "Utilization-Focused Evaluation" (2008) that 'often the

purpose of formative evaluation is to get ready for summative evaluation' (Patton, 2008: 118) and that 'the formative versus summative distinction is context dependent' (idem: 120).

Concerning the judgemental nature of summative evaluation, Scriven is of the view that it is blown out of proportion. He writes that judgement "is not always present in evaluation, which is sometimes merely observational ("He failed to clear the bar at 8'0"; "This joist is seriously flawed by the knotholes"), or mensurational ("She got 92% on multiple-choice test – an easy A according to our well-established standards"). But judgement is usually present, as it is in all interpretative discussions throughout science, and it is there in formative evaluation as well as summative" (Scriven, 1996:154). The statement appears contradictory in so far as judgement is not always present in evaluation but is always present in science, which includes formative and summative evaluations. I suppose the point he is making is that judgement is present in formative evaluation because one has to exercise some judgement in order to know where and what to improve in a programme.

Conceptual vs. instrumental – Scriven uses the distinction Patton used for formative and summative evaluation "Both judgement-oriented (summative) and improvement-oriented (formative) involve the instrumental use of results… Conceptual use of findings, on the other hand, contrasts with instrumental use in that no decision or action is expected" (Patton, 1996:132) to argue what he perceives as an error on the part of Patton that summative evaluation is purely for decision making. He writes:

> "The claim that evaluation is decision-oriented was axiomatic for the Context, Input, Process and Product (CIPP) theorists (its inscription for the book outlining their position). Parenthetically, my first presentation in evaluation (AERA, c. 1968) was mainly aimed to refute that claim. Some examples I used then and often since are evaluations done by historians, which can fundamentally affect our views of history (e.g., evaluating Lincoln's reasons for going to war), or historical figures (evaluating King Edward as a Nazi collaborator)" (Scriven, 1996: 157).

Scriven uses these examples to demonstrate summative evaluation not aimed at decision making. In these cases summative evaluation actually results in knowledge. Therefore, summative evaluation can be used for both research (knowledge support)

and decision making (decision-support) purposes. Formative evaluation is also not solely aimed at improvement. The evaluator has to exercise some judgement in order to realise strengths to build on and weaknesses to improve. As such, formative evaluation also involves some judgement of merit or worth. The difference is a matter of degree – summative evaluation is more 'research-oriented' than 'action-oriented' and the inverse is true of formative evaluation. Patton supports knowledge generation as an independent evaluation purpose by citing syntheses of findings of evaluation research. Scriven acknowledges knowledge generated in this manner, however, he interprets it differently. He argues that the fact that these knowledge generating syntheses are derived from summative findings only serves illustrate the importance of summative evaluation. Scriven perceives knowledge generation as an end product of either summative or formative evaluation, not as an independent purpose.

On developmental evaluation - Scriven contrasted developmental evaluation with distanced evaluation where an evaluator has minimal interaction with programme staff. His perspective is a clear indication of where the two luminaries stand in the paradigms debate which apparently was not over by the time Scriven wrote the article (1996). He responded that "it is clear that the trade-off here is loss of independence, and hence some unknown degree of loss of objectivity and credibility. When the latter considerations are paramount, as is frequently the case, distanced evaluation has the advantage; so the Council's recommendation is not sound in all cases" (p.158). The council referred to is the Council of Foundations which commissioned cluster evaluations that Patton found useful in generating knowledge. Scriven regards developmental evaluation as 'evaluation-related exercises' not evaluation because of the evaluator's involvement in the programmes on a continuous basis. He contends that that makes it an extension of the programme being evaluated. The evaluator actually reviews the programme in a role similar to that of a consultant. Nonetheless, he acknowledges that certain advantages may accrue from such interactions between evaluators and programme staff, for example, educational benefits for staff as well as increased chances of implementation. But the benefits are not sufficient to make 'developmental evaluation' a model for evaluation. He is resolute about the loss of validity that results from prolonged involvement, something which makes the activity less useful. In short, Scriven separates the activities that evaluators get involved in the course of their work (training, advising clarifying *etc.*) from evaluation, which has to include some judgement of merit or worth.

Scriven's description of an evaluator – According to Scriven, a professional evaluator is someone who conducts evaluations. However, he has a very strict conception of evaluation. He defines it as "the process of determining the merit, worth, or significance of things" (Scriven, 1996: 158). While he is not against many activities Patton engages in, in the course of evaluations to improve chances of utilisation (for example, teaching evaluative thinking, advice on the many facets evaluation and so forth); he is steadfast that they are not evaluation. Therefore, people who teach or give guidance on issues relating to evaluation are not necessarily evaluators. He employs the example of a mathematician to illustrate the point:

> "Mathematicians, for example, who are productive members of the mathematics department at a good university, may spend much of their time teaching mathematics, but that doesn't make teaching mathematics part of mathematics. Conversely, being a teacher of mathematics doesn't make someone into a mathematician. Part of what math teachers do in the classroom may be mathematics, part of it may be inspirational history of mathematics. Housekeepers and shopkeepers do mathematics, but that doesn't make them mathematicians, because the usual sense of the term requires that the primary responsibility of a real mathematician's job is doing technically challenging mathematics" (idem: 159).

This means that in order for someone to be called a professional evaluator, they should be able to do 'professionally demanding evaluation as the primary job responsibility'. He further states: "the bottom line here is that someone who can't competently do technically challenging evaluation tasks is lexically excluded from professional status as an evaluator, even if they can and regularly perform the ancillary tasks of a professional" (idem: 159). He then makes a list of ten sets of competencies an evaluator needs to understand and apply (the list is meant to be minimal) which is almost impossible for one evaluator to master. Nevertheless, he dilutes the rigidity of these requirements by accommodating sharing of these technical skills among evaluation teams' members, that is, evaluators in a team may complement one another with these skills in an evaluation. The problem arises when an evaluator is working without a strong technical team. Scriven acknowledges the importance of skills such as interpersonal skills, change management skills and so forth in the conduct of evaluation but cautions against a redefinition of formative evaluation to include them.

The differences between Patton and Scriven stem from their understanding of evaluation. Patton's point of departure is social purposive, where evaluation is understood in terms of its purpose – to alleviate social need – hence his emphasis on utilisation (even though use does not reach the end of the logic model – social improvement). His improvement-orientation tempts him to include certain activities involved in improvement as independent purposes; developmental evaluation is a case in point. Scriven, on the other hand, premises his understanding of evaluation on dictionaries' definitions, as evident in his declaration that his definition is a synthesis of dictionary meanings. Consequently, their views are unlikely to be reconciled. It is useful to find out what activities are involved in developmental evaluation. Patton says they are improvement-oriented but not preparing for summative evaluation. This to me sounds like continuous formative evaluation. Moreover, development is implied in improvement. Knowledge generation can be an independent purpose when an evaluation is designed with it as its goal.

On the surface the gulf between Patton and Scriven manifests itself as a disagreement about the formative-summative distinction; a close scrutiny reveals some deep-seated differences in their perceptions of evaluation. Patton, on the one hand, is fond of collaborative and participatory approaches because of their benefits in enhancing buy-in from stakeholders, which in turn leads to utility. Apart from participation, advice and other activities evaluators help with, these approaches require that stakeholders be trained in the conduct of evaluations so that they can make informed choices. Scriven, on the other hand, is not keen on these approaches because he puts more significance on independence and objectivity more than he values stakeholder involvement. He writes about developmental evaluation that "it is clear that the trade-off here is loss of independence, objectivity and credibility" (Scriven: 1996: 158). For him, collaboration carries the risk of co-option of evaluators and the consequent loss of objectivity which leads to loss of credibility. Scriven is not convinced that the activities involved in collaborative and participatory approaches such as utilization-focused evaluation result in enhanced usefulness of evaluations if their validity is compromised – he believes that is the case with these approaches.

The two luminaries have divergent views on validity and credibility. Patton gives intended users opportunities to determine validity in his call for face validity – the extent to which users understand and believe in data. For him, "credibility includes the

perceived accuracy, fairness, and believability of the evaluation and the evaluator" (Patton, 2008: 396); and information should meet these prescriptions to a satisfactory degree in order for it to be useful. Scriven appears to believe in more predictive forms of validity in the scientific mode by his insistence on independence and objectivity, as well as technical quality (which are necessary for prediction) as corroborated by his definition of an evaluator and the range of skills required. Since technical quality signifies preoccupation with accuracy or unearthing of the truth, Scriven and Patton are on the opposite sides of when it comes to a trade-off between truth and utility. Scriven believes that accuracy, which he perceives to be a result of independence and technical rigour, is a necessary precondition for usefulness. Patton also believes in the importance of truth but he elevates utility above it if a choice is to made due to constraints such as time or resources. His contention is that decision makers prefer some data to no data at all, even if they are less accurate. That is better than very accurate data which is submitted late and therefore, not utilised - truth which serves no one or no purpose.

It is apparent that Scriven values judgement more than improvement as a purpose of evaluation in line with his definition of evaluation – determining the merit, worth or significance of a thing. In this tradition, evaluators do not feel the need to do what Scriven calls evaluation consultation, that is, giving advice and training to stakeholders, clarifying, explaining , follow-up and so forth. Judgement also calls for judges and values to be used in passing judgement. In Scriven's view, evaluators automatically become judges of evaluations by the virtue of their possession of more technical skills than other stakeholders in evaluations, their independence and perceived neutrality. The neutrality derives from the scientific methodology which is perceived to be value-free. This is the position propounded by exponents of the traditional scientific approach to evaluation; and judging by his arguments – overemphasis on independence, objectivity, credibility, validity and technical rigour, Scriven fits comfortably in this group. The issue of neutrality that is supposedly inherent in the scientific methodology raises another question; that of value-pluralism. The claim to be value-free ignores value-pluralism that permeates many societies, particularly democratic ones. In contrast, Patton believes in value-pluralism. Decisions to initiate programmes or interventions are motivated by particular values held by those who want to improve human condition. In most instances, commissioners of evaluations also share the same quests and values and so should evaluators. In a qualitative or quantitative paradigm, choice of instruments, analyses and interpretation is based on a particular value system which is not

necessarily similar to those of intended users and Patton argues that the values of intended users should undergird evaluations.

6.4 Patton challenges Fetterman on empowerment evaluation

"Toward Distinguishing Empowerment Evaluation and Placing it in a Larger Context" (1997) is a review of a book "Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability**"**, edited by David M. Fetterman, Shakeh J. Kaftarian and Abraham Wandersman (1996). Vanderplaat traces the emergence of empowerment evaluation to the emancipatory research that grew out of Paulo Freire's liberation pedagogics. In the book under review, Fetterman places its 'origin in community psychology, action anthropology, the school reform movement and grassroots community development' Patton, 1997: 148). Fetterman *et al*. define empowerment evaluation as "the use of evaluation concepts, techniques, and findings to foster improvement and self-determination" (Fetterman *et al*. [1996] as cited in Patton, 1997a:148).

According to Patton, self-determination is the distinctive feature of empowerment evaluation, which sets it apart from other approaches. However, he wants clarity on some issues which he claims overlap with other approaches. Patton states that empowerment evaluation overlaps with participatory, collaborative, stakeholder-involving evaluations as well as with utilization-focused evaluation. Patton gives five facets of empowerment evaluation as explicated by Fetterman to outline the intersections; they are;

- Training participants to conduct their own evaluations, i.e. capacity-building,
- Evaluators as facilitators and coaches rather than as judges,
- Evaluators advocating on behalf of groups that are disempowered and/or supporting disempowered groups in advocating for themselves,
- Illumination and
- Liberation of those involved (idem: 148).

He suggests, following Cousins (1995) and Cousins, Donohue and Bloom (1995), that participatory and collaborative evaluations vary in terms of goals and 'research-

participants' collaboration'. Research-participants collaboration incorporates the degree of researcher versus practitioner control of the process, depth of participation and breadth of stakeholder participation. Breadth of stakeholder participation refers to the continuum from limited primary users to all legitimate groups. Patton writes:

> "Following these efforts at classification and distinction, Fetterman's overview would suggest that, in addition to these dimensions of participation and collaboration, empowerment evaluation adds attention to and vary along the following continua: (1) the degree to which participants' power of self-determination is enhanced, that is, the extent to which "liberation" occurs; (2) the extent to which evaluators are advocates for disempowered groups or enable groups to advocate for themselves, and (3) the degree to which training participants in evaluation skills is an explicit, primary, and attained outcome of the evaluation process" (Patton, 1997: 149).

Patton contends that since empowerment evaluation interrelates with other types of evaluation with regards to some of the features mentioned, it would be of assistance to elucidate the degrees. For instance, training of participants – to what extent does training empower participants? Does it go beyond that offered in other intersecting evaluations? Does it go far enough to foster self-determination, that is, to enable participants to be able to independently apply what they learned to different environments – what is referred to as 'liberation'? That is empowerment. The degrees make a substantial difference and help in making distinctions between empowerment evaluation and other evaluation approaches with which it is interconnected. Patton asserts that advocacy and training are vehicles to "liberation" which he views as the core of empowerment evaluation. With regards to other facets, that is, facilitation and illumination, he states that they are not unique to empowerment evaluation. However, he is not happy with the degree of emphasis put on self-determination, advocacy and training as an explicit, primary and attained outcome of evaluation. They are not stressed sufficiently in the chapters of the book to distinguish empowerment evaluation from other approaches along the continuum. Patton is not against empowerment evaluation as a distinct approach to evaluation, he is even aware of its distinctive characteristics; what he is not content with is the lack of explicit links in the chapters in the book to these features to demonstrate empowerment evaluation in practice.

Patton asserts that even though there are positive accomplishments in the cited cases, there is no concrete evidence connecting the successes to empowerment evaluation -

not to merely participation and collaboration which it shares with other approaches. Even the claims that might have validated the links between outcomes and EE, it is claimed, are not supported by direct quotations from participants about how they applied the skills they acquired from participating in the processes of empowerment evaluation.

Fetterman's response

Fetterman (1997) acknowledges Patton's contribution to distinguishing empowerment evaluation from participatory and collaborative evaluation. Nonetheless, he is adamant that participation, collaboration, self-determination and capacity building are the characteristics of empowerment evaluation despite their overlaps with other evaluation approaches. He concurs with Patton that empowerment evaluation occupies space along the same continuum with participatory, collaborative, stakeholder-involving and utilization-focused approaches. But he is not in favour of Patton's emphasis on "liberation" as the defining attribute of empowerment evaluation because he deems it limited. He declares:

> "I see a wider range of acceptable adaptations to local circumstances within empowerment evaluation domain. These accommodations to the local context, to participant and evaluator expertise, and to the developmental stage of the program are appropriate and necessary to be effective. It is important to take into account intent and context when determining the boundaries of any effort" (Fetterman, 1997: 255).

He welcomes Patton's proposition for making explicit the degrees of participation along the participation and collaboration continua in order to distinguish empowerment evaluation from other participatory and collaborative approaches to evaluation. He further agrees with Patton that empowerment evaluation is intended for the 'disempowered, disenfranchised, oppressed and economically impoverished populations'. With reference to the chapters in the book, Fetterman is against what he perceives as a parochial perspective by Patton – he claims that Patton employs only one criterion in assessing the chapters in the book, 'faithfulness to empowerment evaluation'. He insists that the chapters emphasise different things like the importance of self-evaluation, the role of quality and so forth; as such, they should not be judged using a single criterion. He further agrees with Patton on the fact the empowerment evaluation has similarities with stakeholder and utilization-focused approaches because 'empowerment evaluation is strongly influenced by these approaches and should

resemble them in many respects' (idem: 256). In response to the advocacy dimension of empowerment evaluation, Fetterman commences with Greene's pronouncement as a way of highlighting the inevitable issue of partisanship that "social program evaluators are inevitably on somebody's side and not on somebody else's side. The sides chosen by evaluators are most importantly expressed in whose questions are addressed and, therefore, what criteria are used to make judgments about program quality" (Greene [1997] as cited in Fetterman, 1997: 257). According to Fetterman, empowerment evaluation is designed to advocate for programmes if the findings of their evaluations are good enough to merit that they be advocated. Negative findings may also be used to advocate for additional resources to help improve programmes.

There is a lot of convergence of opinion between Fetterman and Patton on the loopholes of empowerment evaluation, especially with regard to its distinctiveness to suggest that more work is needed to distinguish it from other participatory and collaborative approaches.

Further debate on distinction of empowerment evaluation

The issue of distinction of empowerment evaluation continues in Patton (2005), a review of *Empowerment Evaluation Principles in Practice* (2005) edited by Fetterman and Wandersman. Patton states that the book is the editors' response to the challenge to differentiate empowerment evaluation from other related evaluation approaches. They do so by 'positing ten principles and showing how they apply in actual cases'. These principles are: "improvement, community ownership, inclusion, democratic participation, social justice, community knowledge, evidence-based strategies, capacity building, organisational learning and accountability" (Patton, 2005: 409). Patton is surprised by the exclusion of self-determination which he believes is the defining characteristic of empowerment evaluation from these principles. Besides the tenacious issue of distinction, Patton is sceptical about the examples cited in the book to illustrate empowerment evaluations in producing the outcomes as stated in the principles because they do not contain evidence of what empowerment evaluation actually achieves. He argues that:

> "These cases epitomize the problems of self-evaluation. The only evidence we have of what these empowerment evaluations accomplished is the testimony from those involved as portrayed

> and interpreted by the evaluators…. No evidence is provided for these assertions" (idem: 410).

What makes Patton more suspicious of the cited examples is because he is aware that Fetterman is well knowledgeable in qualitative methods and techniques for establishing credibility of qualitative data where credibility is always the issue. He argues that Fetterman could have provided direct quotations from participants about how they were affected by empowerment evaluation. Moreover, there is an instance of two respondents answering a question in exactly the same way. This definitely has a negative impact on credibility. Patton writes:

> "The credibility of empowerment evaluation will ultimately rest on evidence that it can deliver on its claims of empowerment, that is, that sustainable capacity is built; that those involved feel a real sense of ownership of the process that makes a difference in how they participate, think, and behave; that organizational learning takes place in a way that makes a difference to organizational performance; that participants become skilled in using evidence and that these skills are generalized and carried forward in practice; that communities can point to the knowledge they have gained and how they have used the knowledge; and that the mutual accountability framework of empowerment evaluation satisfies both internal and external accountability demands" (idem: 411).

Patton feels that these are not proven. He is of the opinion that there is still a lot to be done to verify outcomes of empowerment evaluation with methods like triangulation to ascertain that they are really its intended outcomes. Nevertheless, he acknowledges the efficiency of the book in clarifying empowerment evaluation principles and its outcomes.

Fetterman responds to Patton and Scriven

For the purposes of this section, I shall focus on those elements in Fetterman's response which respond to Patton only and therefore pertinent to the Patton-Fetterman debate. Fetterman's response begins in a rather boastful manner with him recounting the widespread conduct of empowerment evaluation in a selection of countries which represent the five continents of the world. He states that empowerment evaluation is a form of appreciative inquiry to the extent that it is committed to improvement and building on strengths. He then points out what he perceives to be the cause of misstatements about the book *Empowerment Evaluation Principles in Practice* (2005)

on the part of Patton. Fetterman contends that these misstatements stem from Patton's drifting away from the purpose of the book which is "to present the 10 principles of EE and to examine these principles in practice" (Fetterman, 2005: 419). Consequently, Patton assessed the book on the basis of his own purposes; not what the editors intended to accomplish through it, Fetterman argues. Fetterman uses Patton's reference to the U.S.-Iraqi relations to attack him that the latter sees the book as a "political party or movement designed to persuade or dissuade a reluctant if not hostile constituency" (idem: 419). Fetterman paints a contrasting picture of himself as a person concerned about cooperation and advancement of empowerment evaluation.

On the issue of the lack of evidence to support empowerment evaluation's outcomes, Fetterman claims that there is sufficient verification to demonstrate EE's programme outcomes but Patton chose to ignore it. He cites the 'dramatic improvements in academically distressed schools in Arkansas where he provided indices ranging from parenting involvement to discipline' (idem: 419) and the Hewlett Packard Tribal Digital Village project as examples of empowerment evaluation's efficiency in producing desired outcomes – these are the same cases Patton described as 'second-hand summarised testimonies of those involved as portrayed and interpreted by evaluators'(Patton, 2005: 410). He further refers to other unnamed success stories in the book without actually giving direct quotations of what people said.

Patton responds to Fetterman, Wandersman and Snell-Jones

In his response, Patton is lenient with Wandersman and Snell-Jones because they acknowledge the points he raised on the lack of evidence to illustrate the impact of empowerment evaluation. However he does not extend the same goodwill to Fetterman. Patton counters:

> "Those unschooled in appreciative inquiry should not be misled into believing that Fetterman's response represents that approach, which involves more than noting a few areas of agreement before turning defensive and disputatious, all the while denying the same. Labelling the book and his work empowerment evaluation (EE) scholarship and denying its explicit and overt political premises and agenda is disingenuous at best and an unfortunate switch of tactics, as owning the political nature of EE has been one of its strengths" (Patton, 2005b: 429).

While acknowledging the scholarship involved in empowerment evaluation, Patton is concerned about Fetterman's denunciation of political connotations embodied in empowerment evaluation. Patton argues that 'empowerment' is a provocative word in the context of evaluation and that it has political overtones. He writes that "to pretend that such scholarship is non-political and nonadversarial is contrary to the very philosophical and epistemological foundations of EE, or so it seems to me" (idem: 429). He also defends himself against Fetterman's accusation that he strayed from the purpose of the book by lamenting a practice whereby a reviewer is given boundaries within which confine his/her views and criteria to use by book writers or editors when reviewing. This is tantamount to censorship. Patton still feels strongly that there is not sufficient data to link outcomes to empowerment evaluation.

6.5 Kirkhart's dispute with Patton's focus on 'utilisation'

Conceptual restraints inherent in 'use' or 'utilisation'

Kirkhart (2000), asserts that too much time and energy have been expended on efforts to comprehend and foster evaluation use, and Patton's theory on it takes the centre stage. She propounds that the scope of language used in evaluation has hampered an inclusive understanding of evaluation influence. The terms *utilisation* and *use* are narrow because they are associated with data-based influence of findings, and when this parochialism was recognised, other forms of influence were brought under use as an overarching term - process use is one such influence that was co-opted under use, she declares. However, this proved to be the short-term solution and a perpetuation of the construct under-representation that the response was intended to correct, she claims. The reasons put forward are as follows:

> "First, the term use is an awkward, inadequate, and imprecise fit with non-results-based applications, the production of unintended effects, and the gradual emergence of influence over time. Second, when history of influence is traced from the perspective of results-based use, the historical roots of other dimensions of evaluation impact are erased. Process use, for example, incorrectly appears as an afterthought, a late arrival. Third, fitting other types of influence under the results-based paradigm continues to privilege the concept of results-based use" (Kirkhart, 2000: 6).

The author argues that evaluation influence cannot be properly understood if it is viewed primarily from results-based angle with other views constituting its periphery. A reconceptualization in which influence can be examined from multiple vantage

points is required to broaden the context and understanding of evaluation influence and language is critical in this regard. Ironically, she supports this contention against use with the assistance of Patton's pronouncements on language: "The evaluation language we choose and use, consciously or unconsciously, necessarily and inherently shapes perceptions, defines 'reality', and affects mutual understanding" (Patton [2000] as cited in Kirkhart, 2000: 7). She aligns herself with Weiss' assertion that 'utilisation' has an overtone of tools and implements; that it embodies the imagery of instrumental and episodic applications that are inappropriate to evaluation influence, and therefore, should be forsaken. Although she supports Weiss' utterances on utilisation, she argues that Weiss' proposal for a shift to 'use' also falls into the same trap because for Kirkhart, use is as instrumental and as episodic as utilisation. Both terms imply 'purposeful, unidirectional influence' she claims. She states that in order to respond to the question of how and the extent to which evaluation 'shapes, affects, supports and changes' persons and systems, a shift from the parochial use to influence is necessary.

> "A broader construct than use alone is needed to represent this integration – one that does not privilege results-based use over influence stemming from the evaluation enterprise itself, one that does not chronologically limit our vision of the effects of evaluation, one that looks beyond the sight line of our intentions. Toward this end, this chapter purposefully shifts terminology – from use to influence – in proposing an integrated theory" (Kirkhart, 2000: 7).

Influence is defined as the capacity or power of persons or things to produce effects on others by intangible means or indirect means. Kirkhart argues that the term is broader than use and it accommodates the creation of a framework that is suited to examine effects that are 'multidirectional, incremental, unintentional and non-instrumental in addition to those that are covered by use – unidirectional, episodic, intended and instrumental'. These are illustrated in her integrated framework of influence which follows.

An integrated theory of influence

Kirkhart's integrated theory of influence has three dimensions, namely – source of influence, intention and time frame.

Source of influence – This refers to an agent of change which exerts 'power or influence on individuals, organisations or broader decision-making communities'. Sources of influence are two-fold – those influences associated with the process of

evaluation and those associated with results of evaluation. The former source only came into prominence after Patton's discovery of process use.

Results-based influence – The writer avers that early attention to use stemmed from a drive to maximise the positive social impact of evaluation amid concerns of perceived non-utilisation of evaluation findings. Attention focused on the information produced by evaluations and therefore, data-based conclusions were made, she argues. The purposes of evaluations were mostly summative, but later broadened to include formative evaluations, and use was viewed from an instrumental perspective. But it also widened to include other types of use, conceptual and symbolic/persuasion/political uses.

Process-based influence – Kirkhart contends that not all evaluation influence emanates from formative and summative reporting of results. She maintains that sometimes the processes of evaluation can be a major source of influence – process use. She further alludes to the fact that process use did not appear in evaluation literature with Patton's 'discovery' in the 1980s, but that it had been there in change agent roles of the evaluator. She writes: "Process use first emerged in the utilization literature as a means of facilitating results-based use. Greene's focus (1988b), for instance, was on creating conditions conducive to results-based utilization, rather than on the intrinsic effects of the evaluation process itself. Subsequent treatments of process use emphasized its value independent of results-based use (Whitemore, 1991"" (idem: 10). She adopts Greene's three-dimensional framework of process-based influence – cognitive, affective and political. The *cognitive* dimension is related to 'changes in understandings stimulated by the discussion, reflection and problem analysis embedded in the evaluation process'. Sometimes processes may result in instrumental use, but the cognitive dimension is concerned with enhanced understandings among those partaking in the evaluation process. Sometimes being part of an evaluation process arouses feelings of individual or group self-worth and value among those included – this is the *affective* dimension of process-based influence. Apart from the psychological component implied in self-worth, participants develop feelings for the programme/policy and about evaluation itself. The political dimension "addresses the use of the evaluation process itself to create new dialogues, draw attention to social problems, or influence the dynamics of power and privilege embedded in and surrounding the evaluand" (idem: 10). She points to the models that promote evaluation as an 'explicit intervention' to modify

programme operations or to support outcomes as underlining the significance of the political dimension.

Intention – Variously called intention or intentionality, it refers to the extent to which evaluation influence is 'purposefully directed, consciously recognised and planfully anticipated'. It relates to the presence or absence of a purposeful direction to exert influence through evaluation processes or findings. Two types of influence can be discerned – the *intended* influence which is the more visible of the two and is explicit in the purpose of evaluation. *Unintended* influences are those unexpected impacts of evaluation which come in unexpected ways. Intention concerns the intentions of the evaluator, client or other stakeholders in the evaluation and it has three aspects – the type of influence desired or anticipated, who is to be influenced and the persons, processes and findings that are expected to exert influence.

Intended influence – This is the kind of influence that results when an evaluation is purposefully directed to exert influence either through the processes or the results produced. Kirkhart states that Patton's notion of *intended use by intended users* indicates a direct path between intention and influence. Here, potential users of evaluation information are identified early in the evaluation process and their needs drive the evaluation process. Apparent in the definition of intended influence is that it can be results-based and/or process-based. The latter branches out into two streams in participatory evaluation – transformative participatory evaluation and practical participatory evaluation. "In transformative participatory evaluation (T-PE), the intent is empowerment, social action, and change, whereas practical participatory evaluation (P-PE) intends to support program or organizational problem-solving" (idem: 11). The author draws attention to the fact that not all intentions are communicated explicitly – while some may be *manifest*, others may be *latent*. She writes: "However, intended influences could also include latent, covert evaluation functions….For example, an evaluation with the manifest function of improving program effectiveness could also have a latent function of increasing program visibility in the community" (idem: 12). Therefore, intended influence needs attention to both manifest and latent functions as well as understanding the intentions of evaluators, clients and a range of other stakeholder groups. It requires recognition of the plurality of intended uses and intended users.

Unintended influence – This is when evaluation influences programmes and systems in ways that are unanticipated through unforeseen pathways. Kirkhart writes that "attention to the unintended influence of evaluation acknowledges both the power of ripple effects and our inability to anticipate all ramifications of our work. In evaluation, as in the programs themselves, unintended influence may be more impactful than intended influence. Moreover, the territory defined by unintended influence is broader" (idem: 13). Unlike intended influence which focuses on intended uses by intended users, unintended influence has a number of permutations, she argues. One, intended users may exert influence in unintended ways or affect people and organisations other than intended. She illustrates this point with an example whereby the data of an evaluation whose intention to effect changes within the programme, spilled over to the community and galvanised it to form a coalition to advocate for legislative change. Although the influence was intended, unintended influence followed and was broader than the intended one (intended influence by intended users and a further unintended influence by unintended users). The second permutation is the involvement of unintended users in exerting an intended influence. She cites this example:

> "Consider a process-based example in which a needs assessment is conducted on the problem of violence in public schools. The intention was to involve parents and teachers with school board members in identifying concerns and suggesting solutions toward a safe school environment; however, students asserted their interest in the evaluation, and their participation in the needs assessment altered the climate of the school" (idem: 13).

In this instance, the influence was intended (safety) on the intended system (school environment) but it came through an unintended route (students). Thirdly, the users, the nature of influence and systems influenced may all be unintended. For instance, "consider an internal evaluation of a local human service agency intended to support a request for continued funding from its community-based sponsor. As the evaluation unfolded, consumers played an unexpected role in the process, generating unintended positive publicity for the agency. The inclusive evaluation process was cited as a model, and a state wide consumer advocacy group challenged the public sponsor to rethink the parameters of the evaluation that they required for funding" (idem: 13). What followed may have been in addition to the intended use of data for continued funding support.

Time – this refers to the 'chronological or developmental periods in which evaluation influence emerges, exists or continues'. It is subdivided into three – immediate, end-of-cycle and long-term. The writer draws attention to the fact that time is a continuum, and as such, the delineation of periods within time does not presume influence as an event in a certain point in time. It simply points to influence that is visible in given point in time – it may be an event or a process.

Immediate influence – Immediate influence refers to "influence that occurs or is visible concurrent with the evaluation process. Immediate influence may occur during the process of anticipating, planning, and implementing evaluation. It includes early influences that plant seeds of long-term effects or that may show cumulative impact over time as well as short-term effects that may not have long-term ramifications" (idem: 15). It is argued that immediate influence is not limited to the processes of evaluation; it is also results-based. For instance, the results of an evaluability assessment may lay the foundation for adjustments in the course of evaluation. Consequently, 'assessment process represents immediate, intended, results-based influence'. She underscores the point that immediate does not necessarily mean fast paced because it is tied to timeframe of an evaluation. It may evolve slowly when evaluation spans several years, but will be immediate in relation to the point at which it occurs along the continuum. Furthermore, immediate does not refer to duration of influence; it be may immediate influence that is short-lived or that continues beyond the cycle of evaluation and still remain visible.

End-of-cycle influence – This kind of evaluation occurs around the conclusion of a summative evaluation study or of a cycle in a formative evaluation. It includes influences from the products of evaluation (reports, summaries *etc*) and from the processes of dissemination of results. It is asserted that end-of-cycle influence is similar to end of treatment effects in outcome evaluation, drawing attention to the conclusion of an evaluation study or of a given cycle in an ongoing evaluation. "It also includes the process that brings closure to a particular evaluation cycle in the absence of a formal written report and in the context of more developmental use (Patton, 1994,1997)" (idem: 16). Although it comes at the end of a cycle, this influence is not restricted to results; it may also be process-based. For instance, 'networking interactions during evaluation's closure'. "Brett, Hill-Mead, and Wu describe an end-of-cycle influence that bridges process and results when they note that the structured, data-oriented

reflection of the quarterly synthesis process taught staff how to mentor corps members in goal setting for the following year" (idem: 16).

Long-term influence – These are effects that are felt for a long period of time after an evaluation or that evolve over an extended period of time. This kind of influence alerts evaluators to be on the lookout for impacts that are more distant from evaluation chronologically and to follow early impacts over time. Long-term effects may be delayed or long-lasting or both. *Delayed* influence occurs when results, for some reason or the other remain unused. After an extended period of time, their relevance is recognised and they are incorporated into another study or other studies. *Long-lasting* influence is when processes or results exert an ongoing influence from the evaluation process through an extended period after dissemination of results.

The main point of Kirkhart's argument is that 'use' or 'utilisation' gives precedence to intended, immediate and end-of cycle influences in a linear or unidirectional mode and not enough attention to other kinds of influence.

Patton responds to Kirkhart's proposition of influence

Patton notes the reasons Kirkhart prefers influence to utilisation is because influence helps in constructing an integrated theory of evaluation's consequences. It also captures all dimensions of evaluation, namely – source of influence (processes and findings), intention (intended and unintended) and time (immediate, end-of-cycle and long-term). These she argues, capture multidirectional, incremental, unintentional as well as instrumental uses. Patton concedes that her framework is efficient in capturing the above dimensions; however, he argues that it is less useful in informing practice. Referring to Alkin, restates his preference for use:

> "Evaluation use typically refers to the impact of the evaluation (findings and process) within the context of the programme being evaluated, within some reasonable time frame. Evaluation influence refers to the impact on an external program, which may or may not be related to the program evaluated, or to the impact of the evaluation at some future time. An important distinction between evaluation influence and evaluation use is that evaluators who are concerned with evaluation use can actively pursue the course of action to potentially enhance utilization by recognizing the evaluation factors and attempting to be responsive to them, but evaluation influence is more difficult to predict or control"(Alkin [2005] as cited by Patton, 2008:110).

Patton asserts that utilization-focused evaluation is focused on "intended use by intended users". That itself emphasises the centrality of intentionality, therefore, its activities are directed at harnessing that intentionality to enhance utilisation. Kirkhart portrays influence in a similar vein as Weiss by laying emphasis on indirect aspects of evaluation effects over extended time and outside the programme evaluated. This is contrary to Patton's approach to evaluation use which focuses on those aspects of evaluation that can be controlled in order to promote use instead of depending on things that are beyond the evaluator's control.

6.6 Henry's discussion of Patton and the notion of "social betterment"

Henry (2000) seeks to redefine the goal of evaluation which he contends has been displaced by the ascendancy of use. While he understands the circumstances that elevated use – a response to the alarm caused by the realisation that evaluations were not being used directly – and assents to its essentiality, he is concerned that it has now become both the subject of study and a guiding purpose for evaluators, particularly those inspired by Patton's utilization-focused evaluation. He writes:

> "These prescriptions sought to turn weaknesses into strength, making use into the criterion by which the success of an evaluation could be defined. The most fully developed and articulated of these theories, first published in 1978 and significantly refined in later editions, became an influential, if not the most influential, theory of evaluation in the latter part of the century. "Utilization-focused evaluation begins with the premise that evaluations should be judged by their 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"" (Henry, 2000: 85).

He concedes that use is an important consideration in evaluation, but evaluators should not set out with it as the defining goal. He proposes an alternative goal for evaluation, that of *social betterment*, with use serving as a vehicle to it. "The goal of evaluation is not use; the goal is social betterment. Use then, although not the ultimate goal is a means by which evaluation achieves social betterment. However, not all use produces social betterment, and not all forms of social betterment can be achieved by pursuing use. Social betterment means improved social conditions, the reduction of social problems, or the alleviation of human distress" (Henry, 2000: 86). Henry argues that if evaluations are planned with the specific purpose of use in mind, it can evolve a life of

its own, rather than serving as a means to social betterment. This carries the risk of the former displacing the latter as the goal of evaluation and in the process, obscuring the purposes of social betterment – improving social conditions and alleviating human suffering. He further avers that evaluation promotes social betterment indirectly by providing information to those charged with making social policies, those implementing them, stakeholders, consumers and citizens. Promotion of social betterment is indirect because improvement of social conditions is dependent on what is done with information by users, not by evaluators and on the quality of information. He writes:

> "Pursuing social betterment requires that three functions be performed – determining what constitutes the "common good", choosing a course of action leading to the common good, and adapting the chosen course of action to specific circumstances. Each function … requires different information which corresponds to different priorities of evaluation" (idem: 87).

Determining the common good – There is no clear definition of common good, rather, it is defined by what it is not. Henry argues that it is easier for society to reach consensus about what to achieve by confronting its absence and then set out to achieve it than by directly setting out what to achieve because of multiple and conflicting views associated with *value pluralism* that is prevalent in most democratic societies He cites excerpts from several writers to illustrate the point:

> "Human cognitive ability is so flexible and creative that every conceivable moral principle generates opposition and counter-principles (Kendler, 1999: 832). Democracies in which members have the autonomy to express their views are uniquely suited for the generation of oppositional statements, and value pluralism is the inevitable result (Galston, 1999). Defining the common good is therefore a matter of defining a need or a problem around which there is sufficient consensus that it is considered a social problem or, using Berlin's terminology, a human need" (1998)" (idem: 87).

The role of evaluation is to provide empirical information that can support or inform the determination of common good. It may also refute the claim of need or investigate the extent and consequences of a problem. By accentuating the effectiveness of certain social policies, alert and challenge those charged with policymaking and implementation to seek alternative policy solutions.

Selecting a course of action – Evaluations can provide information on whether or not a certain course of action results in social betterment than another. In doing so, it assists with assessing alternative courses of action that can be adopted as policies or in programmes. "For example, Gueron (1997) describes the welfare reform evaluations that tested the effectiveness of work requirements in direct comparison with the existing programs. The reforms resulted in reduced welfare rolls, more employment, and higher wages when compared with the existing system. By providing head-to-head comparison of competing alternatives, these evaluations contributed to the selection of reforms by Congress, the administration, and the state legislatures" (idem: 88). However, Henry cautions that even evaluations that vigorously compare courses of action support rather than supplant making the selection, evaluation information should lead to better decisions by those authorised to make them, instead of evaluators making judgements.

Adapting the course of action – Sometimes a chosen course of action does not yield expected results because of certain weaknesses in implementation or paucity of resources to implement them. Evaluation can help in assessing strengths to build on (assets analysis) and to bring weaknesses to the attention (needs assessment) of those responsible for policy making and to programme directors. This information serves as feedback that enables the adaptation of a chosen course of action in order to achieve better results.

Henry asserts that "To achieve social betterment, collectively we must decide what are the most pressing social problems, decide what actions should be taken to address them, and continue to refine actions that are taken and reengineer the organizations that are charged with achieving the outcomes" (idem: 90). These processes are not necessarily linear, but they may coexist in policy or programme deliberations. Taking cognisance of the limitations of resources, Henry pronounces that these issues cannot be evaluated simultaneously with the same degree of vigour. It is for that reason that questions of prioritisation between utilisation and social betterment become critical. He states that 'a clear focus on social betterment would use an assessment of policy context to allocate evaluation resources'. This means that when there is an outcry from a certain quarter that some social needs have not been addressed, those needs will be assessed (needs assessment to establish common good) and a course of action selected. When the

course of action has been chosen, supported and implemented, an evaluation meant to detect problems with the course of action may be conducted as well as improvement–oriented ones (to adapt the course of action if need be). Therefore, the social betterment option accommodates all the requirements relating to various purposes of evaluation, but that is not the case with use as the focus of evaluation. He writes: "However, with use as a gaol, the perspective narrows. Rather than assessing the policy context, evaluators looking for the most immediate opportunities for use will be drawn to studies that focus on adapting programs or improving organizations" (p. 90). Therefore, focus on use directs resources to evaluations that support the third function required for social betterment, namely adapting the course of action. The author is not averse to use *per se*, but he calls for balance between use and marketing the long-term functions of evaluation.

Patton's response

Patton (2008), acknowledges Henry's proposition to shift emphasis away from use to judging evaluations by their contribution to social betterment only as 'a reminder that use is not an end on itself, inherently good, but must ultimately be judged by its contribution to a better world'(idem: 553). He further concedes ground on the need to move beyond intended uses when he writes that 'social betterment extends the logic model of evaluation beyond short-term and immediate uses to longer-term impacts on society'. However, he is adamant that social betterment cannot take the place of use despite its worthy cause. He states:

> "But it is hard to see how it replaces or supersedes utilisation as a criterion for evaluating evaluations. How can an evaluation contribute to social betterment unless it is used? And social betterment involves a very long time frame. In the short-term, evaluators will still be accountable for intended use by intended users. In so doing, making explicit the assumptions and values that connect intended use by intended users to social betterment is a worthy endeavour (idem: 554).

Patton's response takes the debate back to utilisation versus influence in the sense that it involves an argument about time it takes for social betterment to be realised. This is a question of direct, immediate and intentional use against the long-term influence that eventually results in social betterment.

6.7 Critiques of other evaluation approaches

6.7.1 An assessment of theory-driven approach to validity

In "A Context and Boundaries for a Theory-Driven Approach to Validity" (1989), Patton comments on Chen and Rossi's "The Theory-Driven Approach to Validity" (1987). Patton is not comfortable with what he perceives to be the wide ranging manner in which the theory-driven approach to validity was presented, as if it was applicable to evaluation in general. He argues that the authors also offered the approach as the sole solution to the problems of validity in programme evaluation, yet their definition of evaluation in the original article is parochial as it assumes that the purpose of evaluation is to establish validity both internal and external. Patton further argues that the application of the theory-driven approach is limited. It is narrow because there are many purposes of evaluation and varied stakeholders. Aligned to those purposes and variations are different approaches, and the theory-driven approach to validity is only one of them. Since validity is concerned with internal validity (causality) and external validity (generalisability), Patton argues that Chen and Rossi's approach is concerned with 'generalising causal linkages', and therein comes its limitation. "*However, program evaluations which focus on generating generalizations about causal linkages are a small and narrowly conceptualized subset of all the activity in which evaluators engage*"(Patton, 1989: 375 emphasis original). He asserts that many evaluations are neither concerned with causality nor generalisability since most stakeholders' questions 'are often much more practical and immediate'. Moreover, the rigour of Chen and Rossi's criteria is so inhibitive that they are not applicable to real world situations even where questions of causal linkages and potential generalisability are asked because of their stringency, Patton claims. This kind of approach is said to be applicable to summative evaluations whose aim is to produce generalizations about causal links, but the rigorous criteria needed to establish the cause and effect relationships makes this approach to be less practical in evaluation. "In real practice, under constraints of time, money, and the realities of program implementation, we can seldom provide definitive answers with either causality or generalisability, but we can arrive at some reasonable estimation of the likelihood that particular activities have had particular effects"(idem: 376).

Patton draws a distinction between evaluation researchers and evaluation practitioners. The former are concerned with the establishment of theories and it is necessary to seek out truth, therefore, their employment of rigorous criteria is justified. The latter work in

real world circumstances where things are dynamic and changing and "treatments cannot be controlled across settings enough to establish anything like causal linkages necessary for theory-driven research" (idem: 176). He therefore proposes the 'criterion of reasonableness based on face validity':

> "In this question, given what we know about what the program did, given what we know about the impact of the program, is it reasonable to assume that there was some significant amount of linkage between what the programs did and what resulted?" (idem: 376).

Patton declares that face validity is neither 'prime research criterion' nor 'truth criterion', but an 'action criterion' because it is founded in evaluative action. He states that programme evaluators are sometimes expected to give causal linkages and they must be prepared to do so without recourse to theory-driven approaches to validity. According to Patton, rigour is the criterion for judging research, and the criterion for judging programme evaluation is utility. He contends that theory-driven approach to validity is used mostly by researchers in their quest for truth and is less favourable with practitioners working for specific stakeholders. The preoccupation with theory is premised on the assumption that programmes are undergirded by theory, but there are few clearly conceptualised programmes, he argues. The language of theory-driven approach is a further deterrent to its use because it is so complex that it is incomprehensive to non-researchers. The language problem is accentuated by the fact that it is difficult to get program staff, administrators, and funders to even provide a coherent description of the program in simple, lay terms, Patton claims. He writes:

> "In short, my experience suggests that a program must have achieved an unusual level of clarity and be ready for a major, rigorous summative evaluation before the theory-driven approach to validity will work. Moreover, use of the model requires the level of research sophistication that will be lacking in most stakeholders resulting in the evaluation conceptualization being controlled by sophisticated researchers who understand the model with the consequence that nonresearch stakeholders are left out of the process" (idem: 377).

Patton then proposes an alternative approach which he claims is more widely applicable, and involves working with "primary stakeholders to generate their conceptualization of "the program's theory of action" (idem: 377). In this conceptualisation, non-researchers are involved figuring out what things have to happen

in the programme and their sequence. They construct a means-end hierarchy which includes objectives, activities, inputs and various outcomes; immediate outcomes, middle range outcomes, and long-term impacts. This results in a stakeholder theory of action that is responsive to their information needs from which causal linkages can fathomed. It is different from the Chen and Rossi's 'academically-driven' programme theory which may result in evaluators fitting programmes to models, Patton argues.

6.7.2 Appraisal of feminist evaluation

"Feminist, Yes, but is it Evaluation?" (2002b) is Patton's show of endorsement for feminist evaluation. In the process of doing so and as a point of departure, he initially leans heavily on the book by philosopher Elizabeth Minnich, "Transforming Knowledge" (1990). This will be reflected here because it is the cornerstone of Patton's argument for feminist evaluation against opposition from the establishment. He opens with the meaningfully pregnant and germane words from Minnich that "the root problem reappears in different guises in all fields and throughout the dominant tradition. It is, simply, that the majority of humankind was excluded from education and the making of what has been called knowledge" (Minnich [1990] as cited by Patton, 2002b: 97).

In laying out a major premise and in highlighting the importance of 'concepts' and 'conceptual frameworks' in conveying 'embedded messages about what and who is important', he quotes further; "the dominant few not only defined themselves as the inclusive kind of human but as the norm and the ideal" (Minnich [1990] as cited by Patton, 2002b: 97). The dominant few were men of course, who are reported to have defined themselves as 'mankind/humankind'. " [T]hey created root definitions of what it means to be human that, with the concepts and theories that flowed from and reinforced those definitions, made it difficult to think about, or in the mode of, anyone other than themselves, just as they made it difficult to think honestly about the defining few" (idem: 97). A parallel is drawn between Minnich's perspective on knowledge and programme evaluation which Patton claims in the words of Minnich that "the majority of evaluators were excluded from what has been called knowledge in evaluation. The dominant few not only described themselves as the inclusive kind of evaluator but as the norm and the ideal" (Minnich [1990] as cited by Patton, 2002b: 98). He cites the humiliating ridicule that he endured and resistance while utilization-focused evaluation was still marginalised because of the dominance of the traditional scientific approach.

These were the 'dominant few' who were primarily privileged men extolling the virtues of the quantitative statistics to the exclusion of other approaches, he recalls. The common objection he got was that it was good that UF-E involved intended users to build capacity and so on, but those were dismissed as *not* evaluation. By the same token, he contends, objections to feminist evaluation is the 'echo of the same dismissive and persistent voice' to keep evaluation pure – judge of merit or worth. He writes: "the power to define is the power to control, to include and exclude. That is what Minnich was referring to as the root problem that reappears in different guises in all fields. The root problem is the power of the dominant few to define what constitutes legitimate activity and real knowledge, and these dominant few not only define themselves as inclusive kind of whatever is being defined, in this case evaluator, but also as the norm and the ideal" (Patton, 2002b: 99). The feminist definition of evaluation goes beyond the framework of the traditional perspective evaluation restricted to judging merit and methods because it stresses social justice.

Patton argues that "feminist evaluation can include judging merit or worth, using social science methods to determine effectiveness, and measuring goal attainment, but those are means to a greater end – increased social justice for the oppressed, especially, but not exclusively women" (idem:100). Question marks about the status of feminist evaluation are a consequence of criteria for judging quality and credibility of evaluation founded on particular philosophical underpinnings and purposes for evaluation by the dominant few, making their definitions 'the norm and the ideal'. The feminist evaluation challenges these theoretical orientations and criteria for judging quality and credibility stemming from them.

Patton then comes up with common principles of feminist evaluation gleaned from articles on feminist evaluation (in the New Direction for Evaluation, no. 96, Winter 2002) from which he draws criteria for feminist evaluation. They are:

- Explicitly attending to gender inequalities and , where possible and appropriate, understanding and connecting gender inequalities to other forms of social injustice (racism, social class inequalities)
- Emphasizing the inevitability of the political nature of evaluation (knowledge is power), therefore explicitly addressing power relationships and issues throughout and in all aspects of the evaluation

- Engendering a sense of correctedness and equality between evaluators and those in the programme being evaluated (genuine power sharing)
- Using participatory processes that support consciousness-raising, capacity building and evaluator reflexivity, and make that knowledge and knowledge-creation resources available to those who participate in the evaluation, especially disadvantaged women
- Explicitly acknowledging and valuing "women's ways of knowing", including integrating reason, emotion, intuition, experience, and analytic thought and supporting expression of women's voices about their own experiences and
- Using evaluation processes and findings to bring about change, specifically increased social justice for women and oppressed and disadvantaged people (Patton, 2002b:104).

He then outlines alternative criteria for judging the quality of evaluation from different perspectives within different philosophical frameworks in order to see if they have space for feminist evaluation within them – a factor that is critical in accommodating or rejecting it as evaluation. After having expounded the five sets criteria for judging quality in evaluation from these perspectives, Patton matches feminist evaluation with each of them to locate where it fits. They are:

Traditional scientific research criteria – Some proponents of qualitative inquiry in its early development looked back to the traditional scientific research criteria in an effort to increase legitimacy and credibility of qualitative evaluation. That was a somehow tentative and apologetic approach to qualitative inquiry and was indicative of a yearning to be accepted into mainstream scientific paradigm at that time. Since science emphasised objectivity as a measure against researcher bias, these researchers also stress rigorous and systematic procedures of data collection such as cross-checking and cross-validation of fieldwork, using multiple coding methods to establish validity and reliability of patterns. The purpose is both to explain and describe with maximum accuracy possible, the phenomena as they are in the real world.

Traditional scientific research criteria are criteria employed by the 'dominant privileged few' in the early years of empirical evaluation practice with heavy emphasis on objectivity, credibility, neutrality and truth. According to them, the credibility of evaluation revolves around these principles – and the feminist evaluators' open

advocacy for social justice violates them. They therefore, do not regard feminist evaluation as evaluation.

Social construction and constructivist criteria – This school of thought differentiates between the social and the physical worlds, the former is viewed as a human construction, that is, it is socially, economically and politically constructed, just as human understanding and explanations of the latter are. Advocates of social construction and constructivism have developed certain concepts that are parallel to those employed in traditional scientific research. For instance, Lincoln and Guba put forward "credibility as an analog to internal validity, transferability as an analog to external validity, dependability as an analog to reliability, and confirmability as an analog to objectivity. In combination they viewed these as addressing 'trustworthiness' itself a parallel to the term '*rigor*' (Patton, 2002:267). In tandem with their social constructivist perspective, they seek to report multiple views from multiple constructions rather than a single one. Emphasis is laid on a deep understanding of specific cases within specific contexts as opposed to causal explanations and generalisations of human interactions. Patton declares that:

> "They offer perspective and encourage dialogue among perspectives rather than aiming at a singular truth and linear prediction. Social constructivist's findings are explicitly informed by praxis and reflexivity, that is, understanding how one's own experiences and background affect what one understands and how one acts in the world, including acts of inquiry" (idem: 267-9).

Patton asserts that feminist evaluation can find a home in these criteria because of the value they put on multiple perspectives and multiple truths. However, he intimates that social constructivists may be uncomfortable with the activist inclination of the feminist evaluation evident in its advocacy for social justice. Therefore, social constructivism does not provide enough conditions for the legitimacy of feminist evaluation, he declares.

Artistic and evocative criteria – The previous two sets of criteria for judging qualitative inquiry emphasised the scientific nature of research and a mix of science and art in qualitative studies respectively. This set of criteria in turn, puts stress on

'aesthetics, creativity, interpretive validity, and expressive voice' where findings of case studies are treated as literary works and are performed. Patton writes that:

> "Artistically-oriented qualitative analysts seek to engage those receiving the work, to connect with them, move them, provoke and stimulate. Creative nonfiction and fictional forms of representation blur the boundaries between what is 'real' and what has been created to represent the essence of reality, at least as it is perceived, without a literal representation of that perceived reality" (idem: 269).

Performance of these 'creative syntheses' actually enhances understanding of the analyst's interpretation of findings while simultaneously attaching affective dimension to them. Enlightenment gained, separately or in concert with the emotional/feeling connection engendered by the analyst's expressive voice, may lead to increase interest in and consequent instrumental use.

These criteria are welcoming to feminist evaluation because they draw from the literary traditions of feminism and incorporate them into evaluation, Patton claims. According to him, the greatest influence of feminist literature on feminist evaluation is the latter's focus on voice. He writes: "In addition to encouraging each woman to find her voice, the critical and creative writing involved in feminist literature challenges each woman (really each person) to own her voice and perspective" (Patton, 2002b:104). The feminist critiques of knowledge challenges evaluators to be explicit about ownership of what they write and to accommodate multiple perspectives. This implies acknowledging that biases and limitations on the part of evaluators because of their philosophical postures, which feed their criteria, do not readily permit all perspectives, which is humanly impossible.

Pragmatic and utilitarian criteria – This is the posture taken by Patton in utilization-focused evaluation. An evaluation researcher who assumes this stance is not aligned to any ideological philosophy, but is pragmatism-oriented in the sense that they would employ any model that promotes use. Patton states that the focus of pragmatic utilitarianism is on "answering concrete questions using practical methods and straightforward analysis while appreciating that those who use evaluations apply both 'truth tests' – are the finding accurate and valid? – and 'utility tests' – are the findings relevant and useful?" (Patton, 2002: 271). To achieve these, an evaluator has to match

questions to particular contexts and specific needs of intended users. The evaluator also has to have a wide range of skills in research methods as well as creativity.

These emphasise "the match between the values and perspectives of primary intended users of an evaluation and the evaluator" (Patton, 2002b: 105). Patton states that feminist evaluation is an important option for programmes following feminist principles and for those which are not explicitly feminist but have gender equality as the core of their operations. In such environments, feminist evaluation may be desirable and its criteria match such programme purposes insofar as they advance social justice, in these instances women empowerment to be on par with their male brethren. Patton places utilization-focused evaluation under pragmatic utilitarian criteria and he asserts that feminist evaluation is a viable option for it. Since utilization-focused evaluation stresses a match between the values of intended users and those of evaluators, there is no inherent conflict between it and feminist evaluation as long as there is a match. The intriguing aspect is that Patton writes about feminist evaluation as an option in the utilization-focused evaluation's menu, but the inverse is not contemplated. It appears as if he accepts feminist evaluation in a subservient role to utilization-focused evaluation and not as an equal approach recognised in its own right – that does not sit well with feminists.

Critical change criteria – According to Patton, the concept 'critical' generally connotes 'an interest in and commitment to social change', which makes this set of criteria biased towards social transformation. These criteria are employed by those evaluators who use Critical Theory approaches to fieldwork and their purpose is to expose social, economic and political inequalities. Rather than simply striving for understanding, they criticise the status quo, raise consciousness and endeavour to change the balance of power in favour of the less powerful or the powerless. Their perspective is Marxist to the extent that they are conscious of class differentiations in societies and they seek to make these explicit. "Critical Theory provides both philosophy and methods for approaching research and evaluations as fundamental and explicit manifestations of political praxis (connecting theory and action), and as change-oriented forms of engagement" (Patton, 2002:270). Studies in qualitative evaluation that reflect influence of Critical Theory range from research-oriented approaches whose purpose is to expose injustices to those that actually bring about change. Collaborative and participatory approaches fall under this category because

they build capacity, raise consciousness and prepare stakeholders for action in changing their conditions.

Patton describes the affinity between these criteria and feminist evaluation as follows:

> "The critical change criteria are not only entirely consistent with feminist evaluation, but, indeed, feminist evaluation exemplifies those criteria. Feminist evaluation acknowledges the inevitably political and moral nature of evaluation judgements and challenges evaluators to connect voice and perspective to *praxis* – acting in the world with an appreciation for and recognition of how those actions inherently express social, political, and moral values" (Schwandt [1989] as cited in Patton, 2002b:106).

Critical change criteria are most accommodating to feminist evaluation. They stress the existence of social injustice in society and as such propagate action on behalf of the less powerful by representing their perspectives, exposing injustice and inequalities. By doing so, they encourage evaluators to own their perspectives and report on the plight of the powerless with a view to helping them, just as the feminist criteria do. However, critical change criteria are broader than feminist criteria because while both are concerned with social injustice, the latter lays more emphasis on women whereas the former deals with social injustice in all its manifestations. After all, critical change criteria are meant to be umbrella criteria embracing several approaches that include feminist evaluation.

The pattern which develops from this discussion indicates stringent criteria based on the narrow definition of evaluation as judgement of merit or worth by the traditional scientific approach in the 1960s and the 1970s. Accordingly, feminist evaluation with its preoccupation with advocacy for social justice could not make the grade. The trend moved towards relaxation of criteria most probably because of the intensification of the paradigms debate in 1970s by supporters of the alternative paradigm. That was a struggle to wrest control away from the 'dominant few' social scientific researchers of quantitative predilection by decentralising their power to define evaluation and criteria for judging its quality. Criteria kept on being progressively relaxed and by criteria change criteria, they were flexible enough to provide a home for feminist evaluation.

6.7.3 The review of Egon Guba and Yvonna Lincoln's naturalistic dogmatism

The writings of Patton from the "Alternative Evaluation Research Paradigm" (1975) exhibit his disposition for qualitative/naturalistic evaluation, but he is not averse to the traditional scientific/experimental paradigm with its predilection for quantitative research. He maintains that their differences are not so much an issue of methodological incompatibility as it is about opposite assumptions about the nature of reality. Referring to Guba and Lincoln to illustrate his point he writes:

> "Guba and Lincoln (1981) have argued that the experimentalist (scientific) and naturalistic paradigms contain incompatible assumptions about the inquirer/subject relationship and the nature of truth. The experimental/scientific paradigm assumes that reality is "singular, convergent, and fragmentable", while the naturalistic paradigm holds a view of reality that is "multiple, divergent, and inter-related" (Guba and Lincoln [1981] as cited by Patton, 2002: 464).

In spite of his understanding of the ontological foundations of that Guba and Lincoln stance, the 'rationalist' philosophy of knowledge that pronounces that reality is a social construction of the mind – as opposed to the 'empiricist' view of existence of an objective reality independent of mental constructions, the line of thought that influenced the scientific positivist paradigm – Patton is disturbed about the narrowing of these perceptions of reality to methodological paradigms debate. In arguing for the naturalistic paradigm, Guba and Lincoln are entangled in the same web the traditional experimental paradigm proponents were entrapped in its heydays. Guba and Lincoln write:

> "*Methodologically*, and in consequence of the ontological and epistemological assumptions already made, the naturalist paradigm rejects the controlling, manipulative (experimental) approach that characterizes science and substitutes for it a *hermeneutic/dialectic process* that takes full advantage, and account, of the observer/observed interaction to create a constructed reality that is as informed and sophisticated as it can be made at any particular point in time" (Guba and Lincoln, 1989: 44).

It is the extremism expressed by the authors against the scientific approach that extends to methods that Patton is opposed to. He calls for a pragmatic approach that is inclusive of methods from either set of assumptions when research questions and contexts dictate so. Advice is extended to evaluators not to get caught up in stakeholders' paradigmatic

postures, instead they are encouraged to create conditions of tolerance and focus on information needed by each group. In this way, evaluation is guided by the 'empirical imperative', not by socialisation of stakeholders or evaluators.

This argument is supported by his recounting of an evaluation in which he worked with a group of educators who were divided into two camps, the progressive "open education" group and the traditional "back-to-basics fundamentalists". The former group wanted the employment of qualitative/naturalistic approach while the latter preferred the quantitative/experimental model. He managed to work with both groups by giving each the types of information they wanted, effectively applying multiple methods and measures. He states "My experience suggests that both groups can agree on the evaluation design that includes multiple types of data and that each group will ultimately pay attention to and use "the other group's data"" (Patton, 2008: 465). This is an example of situational responsiveness where an evaluator responds to disparate methodological needs of stakeholders instead of using methodological incompatibility as an excuse not to conduct an evaluation.

Patton treads the middle ground between the two paradigms but slightly to the left (in a case where the scientific paradigm occupies the right extreme of the continuum and the naturalistic paradigm occupies the left edge) because of his emphasis on the primacy of the research problem and context in determining the design. In scientific approach, variables are manipulated to fit in the experimental design. He is at variance with the dogmatic adherence to naturalistic paradigm just as he is against the domination of the scientific paradigm because these postures undermine evaluators' mental flexibility to move between the two perspectives. Moreover, adherence to hard-line stances can only serve either paradigm at the expense of empirical evidence which is likely to be solicited when both approaches are employed.

6.8 Conclusion

These debates with and critiques of other evaluation scholars are very enlightening about the issues involved in evaluation, theories and practice of evaluation. It is important to observe the significance Patton attaches to directness and immediacy of use in contrast to Weiss' indirectness and incrementalism of use at a distant site and future, which is contrary to his utilisation slogan, 'intended use by intended users'. He is inclined towards instrumental use while Weiss is disposed towards conceptual use or

what she terms ‘enlightenment’. We may as well add the planned/unplanned dimension because intended use is planned, while incremental use is unplanned. These differences assume more significance when one considers the fact that by the time this debate took place, Patton had already written about these dimensions of use in Six Honest Serving Men for Evaluation (1985), where he alluded to bias towards direct, immediate and planned use by some evaluators. Therefore, he was aware of the importance of the importance of types of use at the opposite ends of the above dimensions (indirect, incremental and unplanned types of use), but he was deliberately taking a side. This is tantamount to relegating the remaining types of use that are not planned, immediate and direct to bleakness.

Patton’s exchange with Scriven is reminiscent of the rigid positions assumed by researchers and practitioners during the paradigms debate because of Scriven’s insistence on judgement and Patton’s penchant for improvement. Exchanges with Fetterman centred mainly around uniqueness of empowerment evaluation. Kirkhart and Henry made valuable contributions towards development of utilisation theory, the latter extending evaluation purpose further to indicate alleviating human suffering. His critiques of the theory-driven evaluation, the feminist evaluation and Guba and Lincoln’s adherence to naturalistic doctrine are quite elucidating. What stands out from these debates is that Patton is committed to use as the overarching evaluation criterion and he has contributed brilliantly towards its accomplishment.

CHAPTER SEVEN

CONCLUSIONS

The four editions of utilization-focused evaluation represent the major preoccupation of Patton throughout his involvement in evaluation, both as a researcher and practitioner – utilisation of evaluation. This preoccupation is an outcrop of a strongly held belief that for evaluation to contribute towards social betterment, it has to be useful and to be useful it has to be used. The idea was a consequence of influence from several quarters. Most notable among them is the literature in sociology, training in quantitative and qualitative methods, diffusion of innovations and pragmatism.

However, Patton made substantial original contribution to evaluation development. The strengths delineated hereunder demonstrate Patton's creativity in taking evaluation research and practice forward. Through utilization-focused evaluation and its emphasis on utility, Patton has been particularly instrumental in repositioning evaluation practice away from its research orientation which had the effect of making it unresponsive to decision makers' evaluation needs. This almost put an end to the fledgling evaluation profession as manifested by the so-called utilisation crisis in the 1960s and 1970s in the United States of America. The book was published at the time (1978) when evaluation was dominated by research purposes – theory building which necessitated emphasis on internal and external validity (causality and generalizations). Validity was the centre of attention of evaluation researchers/practitioners who mostly employed experimental and quasi-experimental designs. Patton impressed upon evaluators the importance of differentiating between evaluation research and its practice. He appreciated different roles evaluation researchers and practitioners played and their purposes, therefore the need for differences in approaches. It was his unrelenting persistence that an evaluation that is not used is of poor quality, regardless of its methodological rigour and academic acclaim (2001: 169) which helped extricate evaluation practice from the dominance of experimental approach. It also helped place evaluation firmly on the path to recovery of its credibility; usefulness and growth that have seen it become an internationally recognised profession.

Patton's 'discovery' of process use in the 1990s marked a major development in evaluation research and practice. Stakeholder involvement in the processes of evaluation in stakeholder and collaborative evaluations was viewed as a way to increase

chances of utilisation because of the sense of ownership it fosters in them. But there was no full appreciation of the extent of the impact of involvement in the processes of evaluation; and therefore, the processes of evaluation were not harnessed to achieve potential maximum benefits. Patton enlightened evaluators about advantages of purposefully encouraging learning from the processes of evaluations. Since then process use has developed into a major area of study in evaluation. With the development of its typology, the importance of process use has been lifted to the level of findings use. Process use goes a long way towards instilling ownership of evaluations by stakeholders as well as allaying fears. Fear is one of the major causes of non-use of evaluations. The other major contribution Patton made to the typology is on misutilisation. His observation that use is different from misuse and therefore they cannot be part of the same continuum was very insightful and it has resulted in studies in the hitherto understudied area of misutilisation. This is evident in the works of Burry, Alkin and Ruskus (1985), Alkin and Coyle (1988) and Cousins (2004).

Evaluation purposes had been limited to the formative and summative evaluations since Michael Scriven made that distinction in 1967 until Patton proposed additional ones in 1996. Patton was sufficiently insightful to discern the additional purposes which came with growth of the profession, hence his inclusion of knowledge generation and developmental evaluation among others. These purposes have been accepted by the profession of evaluation including Scriven (1996) who acknowledged the knowledge generation purpose of evaluation. Developmental evaluation is also a subject of scholarly articles in evaluation; for instance Ramstad (2009).

Patton's involvement in the so-called paradigms debate was a welcome voice to the qualitative paradigm because of the domination of the quantitative one in educational and social sciences in the 1960s and 1970s. The debate played a significant role in the development of evaluation by elucidating the importance of qualitative approaches. These approaches introduced a differentiation between methods employed in natural sciences and those used in educational and social sciences. In the Alternative Evaluation Research Paradigm (1975), Patton is very outspoken in his defence of the qualitative research while also very forthright in exposing the weaknesses of the quantitative paradigm. His contribution to evaluation through this debate is important because he highlighted gaps in the scientific paradigm and offered ways to close them to promote the utility of evaluations. For instance, he argued against the rigid use of experimental and quasi-experimental designs in almost all contexts as unresponsive to

stakeholder needs. According to him, evaluations should be localized and be responsive to local situations, their treatments and outcomes must be understandable and meaningful to stakeholders. His point of departure is that evaluations should be useful to intended users or else they produce self-serving academic papers. The inclusion of the paradigm of choices reveals the practical orientation of Patton's writings on paradigms as opposed to dogmatic postures assumed by some evaluators on both sides of the paradigms debate. It illustrates that even though he comes out strongly in favour of the alternative paradigm, he still acknowledges the complementarity of the two types of data. A critical factor is the match between a research problem and the research methodology.

As soon as this debate abated, Patton embarked on another mission, to remove the evaluation field from the 'shadows' of other professions and to develop it into a fully fledged profession. He was among the first evaluation scholars to use the space provided by the lull in paradigms debate to concentrate on the issues surrounding the professionalisation of the evaluation field. He envisioned an evaluation profession that is recognised on equal terms with other professions like education, psychology *etc* and given the same respect afforded to them. His vision of evaluation is the one which promotes its utility, quality and credibility. These require sound training in methods and technical skills and true to form, Patton put his hands and his pen where his mouth is by providing learning materials which help with acquisition of these skills. He has written several articles concerned with guidance on the enhancement of quality and credibility of qualitative analysis and how evaluator credibility affects perceptions about evaluation. Patton highlighted the usefulness of case studies in equipping evaluation trainees with situational 'analysis skills, critical thinking and creativity' in addition to the knowledge grounding provided by traditional teaching. As a growing field, evaluation required new methods to meet increased needs necessitated by its growth; and Patton responded by looking beyond evaluation to other professions for useful methods which could be adapted to evaluation to enable it to meet varied demands of disciplines it covers. The other issue in which Patton played a pivotal role with regards to professionalisation was his impression upon evaluators that politics is an integral part of evaluation and that it cannot be avoided. He suggested a pragmatic solution by encouraging evaluators to prepare for politics instead of complaining about political interference in their work.

The vision of accountability as expounded by Patton demonstrates his disposition towards instrumental use which is planned, direct and immediate. This kind of use is inherent in the intentionality of utilization-focused evaluation's emphasis on 'intended use by intended users'. Intentionality – the extent to which evaluation use is 'purposefully directed, consciously recognised and planfully anticipated' (Kirkhart, 2000) plays a central role in Patton's approach. His focus on intended use has certain advantages, most important of which is alignment to the purposes of evaluations. Most evaluations are carried out to accomplish certain goals which are expressed in briefs by evaluation commissioners. These briefs outline intended purposes of evaluations. Therefore, intended uses are direct responses to clearly stated intentions of evaluations. A useful evaluation is the one which attains what it is meant to achieve. Unintended uses are welcome only in addition to the mandate given to accomplish stated intended goals or purposes. The immediacy of use also bodes well for timeframes set by decision makers who normally expect evaluation results earlier than the time required for long-term influence. Direct use is important because it is predictable and an evaluator can follow a course of action with a view to promote evaluation utilisation. Patton is also concerned about the marketability of evaluation that is premised on unplanned, unintended and long-term uses. By their nature, these uses just occur and no one can market their trade based on something they are not sure about. These uses also absolve evaluators from any accountability or responsibility for utilisation of evaluations. Unaccountability for one's performance paints a dim picture of evaluation and can repel potential clients. There is merit in Patton's rationalization of the selection of primary intended users. It is based on the observation that stakeholders often have diverse and competing interests these interests are unlikely to be served by a single evaluation. In order to make an evaluation process workable, an evaluator has to reduce the number of stakeholders; that in turn will narrow the range of possible questions that will focus evaluation.

Utilization-focused evaluation's focus on intended use by intended users is also a limitation. The almost exclusive attention given to primary intended users has certain implications for evaluation. It means that the number of stakeholders is narrowed to an extent that evaluation caters for interests and values of a particular group or two. Further sifting is done on that group through what Patton calls the 'personal factor' in the selection of primary users of evaluation. The 'personal factor' "represents the leadership, interest, enthusiasm, determination, commitment, assertiveness, and caring

of specific, individual people" (2001: 162), he argues. The traits required by the personal factor are not easy to find in individuals, making it difficult to find primary intended users. This list is also not exhaustive – primary intended users have to be teachable in the complexities of evaluation in order for them to participate meaningfully in its processes and to be able to utilise its findings. They should have some education, probably post-matric qualifications. The difficulty is compounded by the fact that there are no rules to follow in determining people who possess these qualities except for evaluator's perspicacity in judging character.

An evaluator has to be very astute and have special skills to be able to recognise individuals with the 'personal factor'. But that is not the end; they have to possess certain characteristics which are not part of methods and technical skills training they receive at universities. He states that:

"These include skills in: (a) communicating effectively (both orally and in writing) with lay people and non-academics; (b) facilitating group decision making and priority-setting processes; (c) clarifying values; (d) resolving and managing conflicts; (e) handling political conflicts; (f) being sensitive to divergent views, cultural differences, and individual values; (g) negotiating, including creating "win-win" outcomes in difficult negotiations; and handling anger, aggression (both active and passive), and hidden agendas" (idem: 164).

It is apparent that very few evaluators meet these preconditions, and Patton asserts that this list 'is far from exhaustive'. By his admission, "not everyone is well-suited by temperament, training, or values to engage in utilization-focused evaluation" (idem: 164). The rarity of its requirements makes it difficult to practice.

The range of skills required from a utilization-focused evaluator has implications for time, labour and money. These preconditions are not for nothing, they mean that an evaluator, on top of providing methods and technical guidance, has to give training of evaluative thinking to stakeholders, negotiate with them, manage conflict, anger and aggression and so forth. These have impact on time evaluations take to complete and costs. The dedication needed from primary users in terms of time and work means that they practically have to work in evaluation as they do in their employment. These factors are prohibitive to widespread employment of utilization-focused approach.

In the light of the difficulties posed by requirements of the 'personal factor' in the selection of participants, it is apparent that the departure of a few primary users poses a major problem with the set-up and they are difficult to replace. The amount of time needed for quality participation does not help in preventing turnover because it is energy sapping. It actually exacerbates the risk of stakeholder turnover.

The strengths of utilization-focused evaluation far outweigh its limitations. One of the greatest strengths of utilization-focused evaluation is its flexibility. Its ability to adapt to multiple and varied contexts (situational responsiveness) allows it to be employed effectively in pursuance of its stated goals – utility and actual use. Patton writes that "utilization-focused evaluation does not advocate any particular content, model, method, theory or even use" (2001: 160). This places it at the intersection of contending perspectives in research and applied social sciences. The implication is that it may be employed by evaluation practitioners irrespective of their paradigmatic affiliation, and most importantly, it permits employment of both quantitative and qualitative designs in a single evaluation approach. This is useful in comprehensively addressing those evaluations which require both kinds of data. This means that it can be successfully employed in developing countries like South Africa. The multiplicity of failed programmes in the country calls for a dedicated approach to their evaluation with unwavering emphasis on utilisation. This can help improve those programmes which falter due to mishaps in implementation or lack of monitoring. It can also serves as a deterrent to unscrupulous elements in the programmes because they will know that their unbecoming behaviour will be uncovered.

BIBLIOGRAPHY

Alkin, M. C. and Christie, C. A. (2002). The use of role-play in teaching evaluation. *American Journal of Evaluation*, 23 (2): 209-218.

Alkin, M. C. and Coyle, K. (1988). Thoughts on evaluation utilization, misutilization and non-utilization. *Studies in Educational Evaluation*, Vol. 14: 331-340.

Anderson, M. et al. (1999). The use of research in local health services agencies. *Social Science & Medicine*, 49: 1007-1019.

Barbie, E. and Mouton, J. (2001). The practice of social research. Oxford University Press Southern Africa.

Brett, B., Hill-Mead, L. & Wu, S. (2000). Perspectives on evaluation use and demand by users: The case of City Year. *New Directions for Evaluation*, 88 (2000) : 71-83.

Bulmer, M. and Warwick, D.P. (1993). *Social research in developing countries: Surveys and censuses in the Third World*. San Francisco, CA: John Wiley & Sons Inc.

Burry, J., Alkin, M. C. and Ruskus, J. (1985). Studies in Educational Evaluation, Vol. 11: 131-157.

Burt, Perrin. (1998). Effective use and misuse of performance measurement. *American Journal of Evaluation*, 19 (3): 1-11.

Caracelli, V. J. (2000). Evaluation use on the threshold of the Twenty-first century. *New Directions for Evaluation*, 88: 99-111.

Caracelli, V. J. (2006). The evaluator-observer comes to the table: A moment of consequential decision making. *American Journal of Evaluation*, 27 (1) : 104-107.

Caracelli, V. J. (1994). Mixed-method evaluation: Developing quality criteria through concept mapping. *Evaluation Practice*, 15 (2): 139-152.

Chelimsky, E. (1977). Improving the use of evaluation: An agency perspective. *The Prison Journal*, 13-18.

Chen, H-T. (1994). Theory-Driven Evaluations: Need, difficulties, and options. *Evaluation Practice*, 15 (1) : 79-82.

Chen, H-T. (1996a). A comprehensive typology for program evaluation. *Evaluation Practice*, 17 (2): 121-130.

Chen, H-T. (1996b). Epilogue: Synthesizing formative/summative evaluation issues and beyond. *Evaluation Practice*, 17 (2): 163-167.

Coffee, A. and Atkinson, P. (1996). *Making sense of qualitative data*. London: Sage.

Cooksy, L. J. (2006). Ethical challenges. *American Journal of Evaluation*, 19: 381-382.

Cooksy, L. J. and Caracelli, V. J. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Evaluation*, 26 (1): 31-42.

Cousins, J. B. (2004). Commentary: Minimizing evaluation misuse as principled practice. *American Journal of Evaluation*, 25 (3) : 391-397.

Cousins, J. B. (1996). Consequences of researcher involvement in participatory evaluation. *Studies in Educational Evaluation*, 22 (1) : 3-27.

Christie, C. A. & Alkin, M. (2003). The user-oriented evaluator's role in formulating a program theory: Using a theory-driven approach. *American Journal of Evaluation*, 24 (3): 373-385.

Davis M. V. (2006). Teaching practical public health evaluation methods. *American Journal of Evaluation*, 27 (2): 247-256.

De Vaus, D. A. (1996). Surveys in social research, 4th ed. London: UCL Press.

Dewey, J. (1916). *Democracy and Education: An introduction to the philosophy of education*. New York: Free Press.

Dewey, J. (1916). *Democracy and Education: An introduction to the philosophy of education* (1966edn.). New York: Free Press.

Dewey, J. (1938). *Experience and Education*. New York: Collier Books

Dey, I. (1993). *Qualitative data analysis*. New York: Routledge.
Denzin, N. K. and Lincoln, Y. S. (1998). *Collecting and interpreting qualitative materials*. London: Sage.

Denzin, N. K. and Lincoln, Y. S. (2000). *Handbook of qualitative research* 2nd ed. London: Sage.

Duffy, B. P. (1994). Use and abuse of internal evaluation. *New Directions for Program Evaluation*, 64: 25-32.

Feinstein, O. N. (2002). Use of evaluations and the evaluation of their use. *Evaluation*, 8 (4): 433-439

Forss, K., Rebien, C. C. & Carlsson, J. (2002). Process use of evaluations: Types of use that precede lessons learned and feedback. *Evaluation*, 8 (1) : 29-45.

Fetterman, D. M. (1997). Empowerment evaluation: A response to Patton and Scriven. *Evaluation practice*, 18 (3): 253-256.

Fetterman, D. M. (2005). In response to Drs. Patton and Scriven. *American Journal of Evaluation*, 26 (3) 418-420.

Friedman, V. J.(2001). Designed blindness: Perspective on program theory evaluation. *American Journal of Evaluation*, 22(2): 161-181.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on Teaching since 1989. *Educational Researcher*, 18 (7): 4-10.

Grasso, P. G. (2003). What makes an evaluation useful? Reflections from experience in large organizations. *American Journal of Evaluation*, 24 (4): 507-514.

Greene, J. C. (2001). Evaluation extrapolations. *American Journal of Evaluation*, 22 (3) : 397-402.

Greene, J. C. (2000). Challenges of practicing deliberative democratic evaluation. *New Direction for Evaluation*, 85: 13-26.

Greer, B. and Mulhern, G. (2002). *Making sense of data and statistics in psychology*. New York: Palgrave.

Grob, G. F. (2006). The evaluator's role in policy development. *New Directions for Evaluation*, 112: 99-108.

Gregory, A. (2000). Problematizing participation: A critical review of approaches to participation in evaluation theory. *Evaluation*, 6 (2): 179-199.

Guba, E. G. (1978). *Toward a methodology of naturalistic inquiry in educational evaluation*. (CSE Monograph Series in evaluation, No.8). Los Angeles: University of California, Center for the Study of Evaluation.

Guba, E. S. & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Hage, J. (1980). *Theories of Organisations: Form, process and transformation*. New York: John Wiley & Sons.

Hammersley, M. (2002). Ethnography and Realism. In A. M. Huberman and M. B. Mites (Eds), *The qualitative researchers companion* (pp.43-56). California: Sage.

Harkness, J. et al. (2003). *Cross-cultural survey methods*. San Francisco: John Wiley & Sons. Inc.

Harnar, M. A. and Preskill, H. (2007). Evaluators' descriptions of process use: An exploratory study. *New Directions for Evaluation*, 116: 27-44.

Henry, G. T. (2000). Why not use? *New Directions for Evaluation*, 88 : 88-98

Henry, G. T. (2003). Influential evaluations. *American Journal of Evaluation*, 24 (4) : 515-524.

Henry, G. T. & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24 (3) : 293-314.

House, E. R. & Howe, K. R. (2000). Deliberative democratic evaluation. *New Directions for Evaluation*, 85: 3-12.

House, E. R. (2003). Stakeholder bias. *New Directions for Evaluation*, 97: 53-56.

Howell, D. C. (1999). *Fundamental statistics for the behavioural sciences (4th edition). A beginner's guide*. Pacific Grove, CA: Duxbury Press.

Huberman, M. and Cox, P. (1980). Evaluation utilization: Building links between action and reflection. *Studies in Educational Evaluation*, Vol. 16: 157-179.

Jacob, S. (2008). Cross-disciplinarization: A new talisman for evaluation? *American Journal of Evaluation*, 29 (2): 175-194.

Johnson, B. J. and Onwuegbuzie. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Research*, 33 (7): 14-26.

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation*, Vol.116: 45-59.

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, Vol. 88: 5-23.

Learmonth, A. M. (2000). Utilizing research in practice and generating evidence from practice. *Health and Education Research*, 15 (6): 743-756.

Leedy, P. D. and Ormrod, J. E. (2005). *Practical research: Planning and design*. Upper Saddle, New Jersey: Pearson Prentice Hall.

Lincoln, Y. Sessions. (1991). The arts and sciences of program evaluation. *Evaluation Practice*, 12 (1): 1-7.

Mark, M. M. & Henry, G. T. (1998). Social programming and policy making: A realist perspective. *New Directions for Evaluation*, Vol. 78: 73-88.

McConney, A., Rudd, A. and Ayres, R. (2002). Getting to the bottom line: A method for synthesizing findings within mixed-method program evaluations. *American Journal of Evaluation*, 23 (2): 121-140.

Michalski, G. V. and Cousins, J. B. (2001). Multiple perspective on training evaluation. *American Journal of Evaluation*, 22 (1): 37-53.

Mark, M. M., Henry, G. T. and Julnes, G. (1999). Toward an integrative framework for evaluation practice. *American Journal of Evaluation*, 20 (2): 177-198.

Mark, M. M. & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10 (1) : 35-57.

Marra, M. (2000). How much does evaluation matter? Some examples of Utilization of the evaluation of the World Bank's anti-corruption activities. *Evaluation*, 6 (1): 22-36.

Melkote, S. R. (2006). Everett M. Rogers and his contributions to the field of communication and social change. *Journal of Creative Communications*, 1 (1): 111-121.

Mohan, R. & Sullivan, K. (2006). Managing the politics of evaluation to achieve impact. *New Directions for Evaluation*, 112: 7-23.

Morabito, S. M. (2002). Evaluator roles and strategies for expanding evaluation process influence. *American Journal of Evaluation*, 23 (3) : 321-330.

Morgan, L. David. (2007). Paradigms lost and paradigms regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1 (1): 48-76.

Morell, J. A. (2005). Why there are unintended consequences of program action, and What are the implications for doing evaluation? *American Journal of Evaluation*, 26(4): 444-463.

Morris, M. and Cohn, R. (1993). Program evaluators and ethical challenges. *Evaluation Review*, 17 (6): 621-642.

Morris, M. and Jacobs, L. R. (2000). You got a problem with that?: Exploring evaluators' disagreements about ethics. *Evaluation Review*, 24 (4): 384-406.

Mouton, J. (1996). *Understanding social research*. Pretoria: J. L. van Schaik.

Mouton, J. (2001). *How to succeed in your Master's and doctoral studies*. Pretoria: Van Schaik Publishers.

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. London: Sage Publications.

Murray, R. (2002). Citizens' control of evaluations: Formulating and assessing alternatives. *Evaluation*, 8 (1): 81-100.

Nelson, M and Eddy, R. M. (2008). Evaluative thinking and action in the classroom. *New Directions for Evaluation*, 117: 37-46.

Oakley, A. (1998). Gender, methodology and people's ways of knowing: Some problems with feminism and the paradigm debate in social science. *Sociology*, 32 (4): 707-731.

Patton, M. Q. (1975). *Alternative evaluation paradigms*. New York, N.Y: Rockfeller Bros.

Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverly Hills, CA: Sage.

Patton, M. Q. (1980). Making methods choices. *Evaluation and Program Planning*, Vol. 3 : 219-228.

Patton, M. Q. (1980). Truth or consequences in evaluation. *Teaching Evaluation*, 1 (3): 261–283.

Patton, M. Q. (1982). [Review of the book "Metaphors for evaluation: Sources of new methods" by Nick Smith]. *American Journal of Evaluation*, Vol. 3: 48-52.

Patton, M. Q. (1983). Designing evaluations of educational and social programs. *American Journal of Evaluation*, Vol. 4 : 29 – 33.

Patton, M. Q. (1983). Similarities of Extension and evaluation. *Journal of Extension*, 14-21.

Patton, M. Q. (1984). [Review of the book "Beyond method: Strategies for social research" by Gareth Morgan, Ed.]. *American Journal of Evaluation*, Vol. 5: 70-73.

Patton, M. Q. (1984). An alternative approach for problem-solving training program: A utilization-focused evaluation process. *American Journal of Evaluation*, Vol. 7: 189-192.

Patton, M. Q. (1985). *Six honest serving men for evaluation*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Patton, M. Q. (1985). Logical incompabilities and pragmatism. *Evaluation and Program Planning*, 8 (4): 307–308.

Patton, M. Q. (1986). *Utilization-focused evaluation (2nd Ed)*. Thousand Oaks, CA: Sage.

Patton, M. Q. (1986). [Review of the book "Decision-oriented educational research" by William Cooley and William Bickel]. *American Journal of Evaluation*, 7: 71-74

Patton, M. Q. (1987). The mission of the American Evaluation Association. *American Journal of Evaluation*, 8 : 58–60.

Patton, M. Q. (1987). Evaluation's political inherency: Practical implications for design and use. In *Politics of program evaluation, yearbook* (Vol. 15: 100-143).

Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage.

Patton, M., Q. (1987). Evaluation's political inherency: Practical implications for design and use. In D. J. Palumbo (Ed.), *Politics of Program Evaluation*. Newbury Park, CA: Sage.

Patton, M Q. (1988). The evaluator's responsibility for utilization. *Evaluation Practice*, 9 (2): 5–24.

Patton, M. Q. (1988). The future and evaluation. *American Journal Of evaluation*, 9: 90-93.

Patton, M. Q. (1988). Politics and evaluation. *American Journal of Evaluation*, 9: 89-93.

Patton, M. Q. (1988). Evaluator's responsibility for utilization. *Evaluation Practice*, 9 (2): 5-24.

Patton, M. Q. (1988). How primary is your identity as an evaluator? *Evaluation Practice*, 9 (2): 87-92.

Patton, M. Q. (1988). Extension's future: Beyond technology transfer. *Science Communication*, 9 (4): 476-491.

Patton, M. Q. (1989). A context and boundaries for a theory-driven approach to validity. *Evaluation and Program Planning*, Vol. 12: 375-377.

Patton, M. Q. (1989). The Sage difference. *American Journal of Evaluation*, 10: 87 – 89.

Patton, M. Q. (1990). The challenge of being a profession. *American Journal of Evaluation*, 11 (1): 45–51.

Patton, M. Q. (1991). Towards utility in reviews of multivocal literatures. *Review of Educational Research*, 61 (3): 287–292.

Patton, M. Q. (1991). Letters: Unmoved but not unchanged. *Educational Researcher*, 20: 33.

Patton, M. Q. (1993). Developmental Evaluation. *Evaluation Practice*, 15 (3): 311–319.

Patton, M. Q. (1994). Programme Evaluation Standards. *Educational and Psychological Measurement*,

Patton, M. Q. (1994). The program evaluation standards: How to assess programs of educational programs by the Joint Committee for Standards for educational evaluation. *Evaluation Practice*, 15 (2) 193-199.

Patton, M. Q. (1996). A world larger than formative and summative. *Evaluation Practice*, 17 (2): 131–144.

Patton, M. Q. (1997). *Utilization-focused evaluation*: *The twentieth century text (3rd Ed)*. Thousand Oaks, CA: Sage.

Patton, M. Q. (1997). Toward distinguishing empowerment evaluation and placing it in a larger context [Review of the book "Empowerment evaluation: Knowledge and tools for self-assessment" by David Fetterman, Shakeh J. Kaftarian and Abraham Wandersman, Eds]. *Evaluation Practice*, 18 (2): 147 – 163.

Patton, M. Q. (1997). Of vacuum cleaners and toolboxes: A response to Fetterman's response. *Evaluation Practice*, 18 (3): 267-270.

Patton, M. Q. (1998). The challenges of diversity in evaluation: Narrow versus expansive perspectives. *Science Communication*, 20 (1): 148-164.

Patton, M. Q. (1998). Discovering process use. *Evaluation,* 4 (2): 225–233.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative data. *Health Services Research*, 34 (5): 1189– 1208.

Patton, M. Q. (1999). Myths as normative frames for qualitative interpretation of life stories. *Qualitative Inquiry*, 5 (3): 338-352.

Patton, M. Q. (1999). [Review of the book "Realistic evaluation" by Ray Pawson and Nick Tilley]. American Journal of Evaluation, 20 (2): 385-388.

Patton, M. Q. (1999). Comment on Agar: How long will the rose of complexity theory bloom? *Field Methods*, 11 (2): 121-123.

Patton, M. Q. (2001). Use a criterion of quality in evaluation. *Advances in Program Evaluation*, Vol.7: 155 – 180.

Patton, M. Q. (2002). Two decades of developments in qualitative inquiry: A personal experiential perspective. *Qualitative Social Work*, 1 (3): 261–283.

Patton, M. Q. (2002). Teaching and training with metaphors. *American Journal of Evaluation*, 23 (1): 93-98.

Patton, M. Q. (2002). [Review of the book "Learning in the field: An introduction to qualitative research" by Gareth B. Rossman and Sharon F. Rallis, Eds]. *American Journal of Evaluation*, 23 (1): 115-116.

Patton, M. Q. (2002). A vision of evaluation that strengthens democracy. *Evaluation*, 8 (1): 125–139.

Patton, M. Q. (2002). Evaluation, knowledge management, best practices and high quality lessons learned. *American Journal of Evaluation*, 22 (3): 32–336.

Patton, M. Q. (2002). Feminist, yes, but is it evaluation? *New Directions for Evaluation*, No.96: 97-108.

Patton, M. Q. (2002). Weiss' call for humility: Further reflections. *American Journal of Evaluation*, 23 (6): 231–233.

Patton, M. Q. (2003). Inquiry into appreciative evaluation. *New Directions for Evaluation*, Vol. 100: 85–98.

Patton, M. Q. (2003). Distant echo of a strong voice: Reflections on Margaret Mead's evaluation of the Salzburg Seminar. *American Journal of Evaluation*, 24 (1): 123–131.

Patton, M. Q. (2004). The roots of utilization-focused evaluation. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists views and influences* (pp. 276-293). California: Sage.

Patton, M. Q. (2004). A microcosm of the global challenges facing the field: Commentary on HIV/AIDS monitoring and evaluation. *New Directions for Evaluation*, No.103: 163-171.

Patton, M. Q. (2005). The challenges of making evaluation useful. *Ensiao:aval.pol.publ.Educ*., 13 (46): 67-78.

Patton, M. Q. & Patrizi, P. (2005). Case teaching and evaluation. *New Directions for Evaluation*, No.105: 5-14.

Patton, M. Q. (2005). Diverse and creative uses of cases for teaching. *New Directions for evaluation*, No.105: 91-100.

Patton, M. Q. (2005). Towards distinguishing empowerment evaluation and placing it in a larger context: Take two. [Review of the book "Empowerment evaluation principles in practice" by David Fetterman and Abraham Wandersman (Eds]. *American Journal of evaluation*, 26 (3): 408-414.

Patton, M. Q. (2005). Patton responds to Fetterman, Wandersman and Snell-Johns. *American Journal of Evaluation*, 26 (3): 429-410.

Patton, M Q. (2005). Patton responds to Fetterman, Wandersman and Snell-Johns. *American Journal of evaluation*, 26 (1): 429–430.

Patton, M., Q. & Patrizi, P. (Eds). (2005). Teaching evaluation using the Case Method: *New Directions for Evaluation*. San Francisco, CA: Jossey-Bass.

Patton, M. Q., Campbell, M. S. & Patrizi, P. (2005). Evaluation of the Central Valley Partnership of the James Irvine Foundation, *New Directions for Evaluation*, No.105: 39–57.

Patton, M. Q. (2007). Process use as Usefulism. New Directions for Evaluation, 116: 99-112.

Patton, M. Q. (2008). *Utilization-focused evaluation (4th ed)*. Thousand Oaks, CA: Sage.

Pawson, R. and Tilley, N. (1998). Caring communities, paradigm polemics, and design debates. *Evaluation*, 4 (1): 73-90.

Pollitt, C. (1999). Stunted by stakeholders? Limits to collaborative evaluation. *Public Policy and Administration*, 14 (2): 77-90.

Preskill, H. & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, 88 (2000): 25-37.

Preskill, H., Zuckerman, B. & Matthews, B. (2003). An exploratory study of process use: Findings and implications for future research. *American Journal of Evaluation*, 24 (4) : 423-442.

Redwood, S. (2005). Colliding discourses: Deconstructing the process of seeking ethical approval for a participatory evaluation project. *Journal of research in Nursing*, 10(2): 217-230.

Reynolds, A. (1998). Confirmatory program evaluation: A method for strengthening causal inference. *American Journal of Evaluation*, 19 (2): 202-221.

Rogers, E. M., Singhal, A. and Quinlan, M. M. (2004). Diffusion of Innovations. In D. Stacks and M. Salwen (Eds). *An integrated approach to communication theory and research*. New York: Routledge.

Rosas, S. R. (2006). Nonparticipant to participant: A methodological perspective on evaluator ethics. *American Journal of Evaluation*, 27 (1) : 98-193.

Rossi, P. H. (2007). Formative for program evaluation. *American Journal of Evaluation*, 28 (2): 199-202.

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004). *Evaluation: A systematic approach.* Thousand Oaks, CA: Sage Publications.

Rossman, G. B. & Rallis, S. F. (2000). Critical inquiry and use as action. *New Directions for Evaluation*, 88: 55-69.

Russel, M. and Mugyenyi, M. (1997). Armchair empiricism: A reassessment of data collection in survey research in Africa. *African Sociological Review*, 1 (1): 16-29.

Ryan, K. et al. (1998). Advantages and challenges of using inclusive evaluation approaches in evaluation practice. *American Journal of Evaluation*, 19 (1): 101-122.

Schmitt, L. N. T. and Whitsett, M. D. (2008). Using evaluation data to strike a balance between stakeholders and accountability systems. *New Directions for Evaluation*, 117: 47-58.

Scriven, M. (1991). Beyond formative and summative evaluation. In M.W. McLaughlin and D. C. Phillips (Eds), *Evaluation and Education: At Quarter Century*, 39-83

Scriven, M. (1994). The fine line between evaluation and explanation. *Evaluation Practice*, 15 (1): 75-77.

Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice*, 17 (2): 151-161

Scriven, M. (2005). [Review of the book Empowerment evaluation principles in practice by David M. Fetterman and Abraham Wandersman, (Eds)]. *American Journal of Evaluation*, 26 (3): 415-417.

Sengupta, S. (2002). Commentary: begin with a good program theory: The case of the missing guiding principle. *American Journal of Evaluation*, 23 (1): 103-106.

Shadish, W. R. & Epstein, R. (1987). Patterns of evaluation practice among members of the Evaluation Research Society and Evaluation Network. *Evaluation Review*, 11 (5): 555-590.

Shulla, L. M. & Cousins, J. B. (1986). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18 (3): 195-208.

Simons, H. (2004). Utilizing evaluation evidence to enhance professional practice. *Evaluation*, 10 (4): 410-429.

Smith, N.L. (1994). Clarifying and expanding the application of program theory-driven evaluations. *Evaluation Practice*, 15 (1): 83-87.

Smith, N. L. & Chirop, S. (1989). The Weiss-Patton debate: Illumination of fundamental concerns. *American Journal of Evaluation*, 10 (1) : 5-13.

Sutherland, S. (2004). Creating a culture of use for continuous improvement: A case study of Edison Project School. *American Journal of Evaluation*, 25 (3): 277-293.

Telfair, J. (1999). Improving the prospects for a successful relationship between community and evaluator. *New Directions for Evaluation*, 83 (1999) : 55-65.

Telfair, J. & Leviton, L. C. (1999). The community as client: Improving the prospects of useful evaluation findings. *New Directions for Evaluation*, 83 (1999) : 5-14.

Tesch, R. (1992). Qualitative research: *Analysis types and software tools*. Hampshire: The Falmer Press.

Thurston, W. E. & Potvin, L. (2003). Evaluability Assessment: A tool for incorporating evaluation in social change programmes. *Evaluation*, 9 (4) : 453-469.

Torres, R. T., Preskill, H. & Piontek, M. E. (1997). Communicating and reporting: practices and concerns of internal and external evaluators. *Evaluation Practice*, 18 (2) ; 105-125.

Torres, R. T. & Preskill, H. (1999). Ethical dimensions of stakeholder participation and evaluation use. *New Directions for Evaluation*, 82 (1999) : 57-66.

Torres, R. T. & Preskill, R. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, 22 (3): 387-395.

Toulemonde, J. (1995). Should evaluation be freed from its causal links? An answer illustrated by European Economic Development Policies. *Evaluation and Program Planning*, 18 (2): 179-190.

Van der Knaap, P. (2004). Theory-based evaluation and learning: Possibilities and challenges. *Evaluation*, 10 (1): 16-34.

Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, 8 (1): 60-80.

Wallerstein, N. (1999). Power between evaluator and community: research relationships within New Mexico's healthier communities. *Social Science & Medicine*, 49: 39-53.

Weiss, C. H. (1988). Evaluation for decisions: *Is anybody there*? *Does anybody care*?. *Evaluation Practice*, 9 (1) : 5 -19.

Weiss, C. H. (1988). If program decisions hinged only on information: A response to Patton. *Evaluation Practice*, 9 (2) : 87-92.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19 (1): 21-35.

Weiss, C. H. (1999). The interface between evaluation and public policy. *Evaluation*, 5 (4): 468-486.

Weiss, C. H. (2002). Patton's (2001) Lessons learned": A comment. *American Journal of Evaluation*, 23 (2): 229-230.

Weiss, C. H., Murphy-Graham, E. & Birkeland, S. (2005). An alternative route to policy influence: How evaluations affect D.A.R.E. *American Journal of Evaluation*, 26 (1): 12-30.