METAEVALUATION OF WORLD BANK EVALUATIONS OF POVERTY REDUCTION INTERVENTIONS IN UGANDA

A thesis submitted to the
Faculty of Commerce, Law and Management at the University of the Witwatersrand, Johannesburg in fulfilment of the requirements for the Degree of Doctor of Philosophy.

Peter Misigalo Wandwasi

Student Number: 514648

29 November 2017

ABSTRACT

Metaevaluation (evaluation of evaluations) contributes significantly in clarifying and resolving evident causes of inconsistencies in the evaluation outcomes, and enhances the effectiveness of development interventions. It is a specialised form of evaluation, which helps in judging the strengths and weaknesses of a primary evaluation so that interested parties in the intervention are able to judge the reliability and credibility of an evaluation and the evidence used to make decisions. The problem this research sought to solve was that, whilst it was evident that the majority of poverty-reduction interventions funded by the World Bank through its Poverty Reduction Support Credits (PRSCs) in Uganda had been systematically evaluated in order to measure their effectiveness, there was overall evidence of remarkable inconsistencies in the evaluation outcomes of the effectiveness of these poverty-reduction interventions in Uganda.

Using four selected evaluation cases of poverty-reduction interventions of the World Bank, funded through its PRSCs and mixed methods approach (case studies, document analysis and descriptive frequency statistics) based on the Program Evaluation Standards (PES) developed by the Joint Committee in 1994, the main task of the research was to assess how effectively these selected evaluations met the requirements of a good evaluation in order to clarify and resolve evident causes of such inconsistencies. The theoretical basis for good programme evaluation upon which this thesis was based is driven mainly by the quest to promote accountability, transparency and social enquiry, and heightens the demand for increased use of participatory approaches in designing and evaluating development interventions. However, within the context of this theoretical basis, the existing compulsory associated standards for evaluating the quality of evaluations based on the PES provide limited methodological scope and coverage towards assessing the quality of evaluations of social development interventions, which exacerbated these inconsistencies in evaluation outcomes of the effectiveness of poverty-reduction interventions in Uganda.

Confronted with this conundrum, the researcher makes one knowledge contribution to the methodology for conducting metaevaluations based on PES. Contrary to heightened demand for increased use of participatory approaches in designing and evaluating social development interventions based on the theoretical basis, which underpinned this study, the evidence from the analysis of the findings surprisingly revealed that a full range of stakeholders was not adequately identified and engaged in the designing and evaluation processes. Whilst stakeholder identification is not prescribed as an exclusive compulsory associated standard based on the PES by which to measure the quality of evaluations, it is empirically plausible to elevate stakeholder identification grounded in the interactive style of evaluative inquiry which is based on the assumption that those with direct vested interests in the intervention should also control the evaluation of the intervention as an additional independent

compulsory associated standard under Utility Standards based on the PES when conducting metaevaluations of evaluations of not only poverty-reduction interventions, but also other social development interventions.

The elevation of Stakeholder Identification (US1) as an additional independent compulsory associated standard under Utility Standards (US) based on the PES can play an important role of not only contributing towards a better resolution of plausible causes of inconsistencies in the evaluation outcomes of the effectiveness of social development interventions, but can also contribute to a better understanding of how, in practice, the identification and engagement of a full range of stakeholders enhances the promotion of the principles of accountability, transparency and participation, all of which are grounded in theories for good programme evaluation. These principles, if correctly applied, can also insulate evaluation processes of social development interventions from other causes of inconsistencies, such as bureaucratic bottlenecks and conflict of interest.

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX TABLES

DECLARATION

I declare that this thesis, unless otherwise where acknowledged, is my own, unaided work. It is submitted in fulfilment of the requirements of the degree of Doctor of Philosophy in Monitoring and Evaluation with specialised focus on Metaevaluation at the Wits School of Governance in the Faculty of Commerce, Law and Management, University of the Witwatersrand, Johannesburg and has not been submitted before for any degree or examination in any other university.

Peter Misigalo Wandwasi

ACKNOWLEDGEMENT

I am very grateful to Prof. Pundy Pillay for his partial supervisory role he played in reading and commenting on my scripts during preliminary stages of my PhD defence process. I thank Prof Michael Muller for his meticulous supervisory part he played in reading, commenting and providing guidance when carrying out quality assurance corrections during the post-examination period of this PhD thesis. I gratefully acknowledge the invaluable support provided by Dr. Gillian Godsell in reading and providing candid comments on the quality of my writing style.

I thank Prof. Patricia Rogers, Professor of Public Sector Evaluation at RMIT University in Australia for recognising the usefulness of this thesis and recommending its use by her doctorate student in evaluation. Prof. Rogers' assessment of the usefulness of this thesis to her doctoral student in evaluation motivated me to improve further on its robustness. I am also grateful to Dr John Mayne from Canada, an expert in public sector performance evaluation, who is also considered to be the originator of the concept of Contribution Analysis (CA) in the discipline of evaluation, for providing me with useful comments on the problematisation of the research problem in my literature review, which contributed towards improving the robustness with which I framed the problem statement.

I also significantly benefited from various presentations I made at the PhD forum workshops organised by Dr. Horacio Zandamela of Wits School of Governance (which I always made sure I do not miss) from where, I received useful comments, which improved the quality of this PhD thesis.

I thank Prof. Imraan Valodia (Dean of the Faculty of Commerce, Law and Management, University of the Witwatersrand), Prof. David Everatt (Head of Wits School of Governance), Prof. Anthoni Van Nieuwkerk (former Academic Director, Wits School of Governance) and Prof. Patrick Fitzgerald (former Director for Executive Education, Wits School of Governance) for their steady confidence in my unwavering aptitude to succeed which, served as a source of additional oxygen, thus, enabling me to stay resilient in the midst of contrived sources of excruciating distress and waning hope. I also thank Julia Brown from plainenglishproofreading.co.uk based in the United Kingdom for proofreading this thesis which contributed in improving its clarity in language style.

I will always be grateful for God's unconditional protection and gracious provision, without which I could not have been able to have the gift of life and resources to reach this level of achievement.

DEDICATION

This thesis is dedicated to the following people:
My beloved late father, Festo Mabonga (R.I.P),
My late great friend, cousin sister and adopted mother, Edisa Nabalayo (RIP),
My Mentor at the United Nations Joint Inspection Unit of the United Nations System, Ms Susanne Frueh and my friend, Ms Marlene Laeubli-Loud.

To my son, Theo Festo Mabonga Wandwasi.

"Success Is When You Don't Give Up."

LIST OF ACRONYMS

AAP: African Action Plan
AEA: American Evaluation Association
AfrEA: African Evaluation Association
AS1-AS12: Accuracy standards from 1 to 12
CAS: Country Assistance Strategy
CA: Contribution Analysis
DEGEVAL: German Evaluation Society
DRC: Democratic Republic of Congo
FS1-FS3: Feasibility Standards from 1 to 3
GDP: Gross Domestic Product
GNP: Gross National Product
HIPC: Heavily Indebted Poor Countries
ICESCR: International Covenant on Economic, Social and Cultural Rights
NRA: National Resistance Army
NRM: National Resistance Movement
ODA: Overseas Development Association
OECD: Organization for Economic Cooperation and Development
PEAP: Poverty Eradication Action Plan
PES: Program Evaluation Standards
PMB: Produce marketing Board
PRSCs: Poverty Reduction Support Credits
PRSP: Poverty Reduction Strategy Paper
PS1-PS8: Propriety Standards from 1 to 8
SEVAL: Swiss Evaluation Society
UBS: Uganda Bureau of Statistics
UN: United Nations
UNCTAD: United Nations Conference on Trade and Development
UNDP: United Nations Development Programme
UNGA: United Nations General Assembly
US1-US7: Utility Standards from 1 to 7
USSR: Union of the Soviet Socialist Republics

CHAPTER 1: SUMMARISED VERSION OF THE THESIS

1.0 Introduction

This study entitled "Metaevaluation of World Bank evaluations of poverty-reduction interventions in Uganda" evaluated the quality of selected evaluations of poverty reduction interventions using a mixed methods research approach composed of case studies, document analysis and descriptive frequency statistics based on the evaluation quality checklist prescribed under the Program Evaluations Standards (PES).

The structure of this summarised version of the research is composed of the following chapters: Chapter One (Substantive introductory discussion), Chapter Two (Literature Review: Summary Provided), Chapter Three (Research Methodology: Summary Provided), Chapter Four (Presentation of findings: Summary Provided); Chapter Five (Analysis of the findings and answers to research question: Summary Provided) and Chapter Six (Conclusions, contribution to knowledge, suggested areas for further research: Summary Provided).

1.1 Chapter One: Introduction

This chapter makes the case for metaevaluation and advances the plausible perspective that metaevaluation helps in judging the strengths and weaknesses of a primary evaluation, so that, interested parties can be able to judge the reliability and credibility of evaluations and the evidence used to make decisions.

Centered on the principles of accountability, transparency and participatory social enquiry which are grounded in the theory for good programme evaluation, metaevaluation improves both the quality of evaluations as well as the effectiveness of development interventions. Metaevaluation also contributes significantly in clarifying and resolving evident causes of inconsistencies in the evaluation outcomes, and enhances the effectiveness of development interventions.

In this chapter, the problem statement (remarkable inconsistencies in the evaluation outcomes) is specified; the knowledge gap (scanty evidence in the literature on the utilisation of metaevaluation in Uganda, which identifies and resolves the causes of inconsistencies in the evaluation outcomes of the effectiveness of development interventions) is stated; the purpose (to evaluate the quality of selected evaluations so that evident causes of such remarkable inconsistencies in evaluation outcomes could be clarified and resolved) is outlined; the research question (how effectively did selected evaluations conducted under the auspices of the World Bank on poverty reduction interventions funded through the PRSCs Give abbreviation in full on first use in Uganda meet the requirements of a good evaluation?) is also stated.

1.1.1 The Case for Metaevaluation

Metaevaluation is defined by Scriven (1969) as the evaluation of evaluations. Metaevaluation is a unique and specialised form of evaluation, which helps in judging the strengths and weaknesses of a primary evaluation so that interested parties are able to judge the reliability and credibility of evaluations and the evidence used to make decisions (Stufflebeam, 2000; Stufflebeam, 2001; Stufflebeam and Shinkfield, 2007; Cooksy and Caracelli, 2009).

Since metaevaluation is a form of evaluation, there is a variety of definitions of the term ‘evaluation’ reflecting different perspectives of various theorists regarding the purpose and functions of an evaluation (Tyler, 1942, Tyler, 1950; Cronbach, 1963; Alkin, 1969; Alkin, 1970; Scriven, 1969; Glass, 1969; Wittrock and Wiley 1970; Stufflebeam, 1971; Nevo, 1983; Borich, 1990). For example, whilst Tyler (1942) propagates the perspective that evaluation is a process of determining the extent to which the objectives of an intervention are realised, Cronbach (1963), Alkin (1969) and Stufflebeam (1971) assert that evaluation is a process for providing evidence for decision-making.

The diverse definitions of the purpose of evaluation demonstrate that the field of evaluation continues to attract different intellectual orientations. Whilst there are many different theoretical perspectives regarding the purpose of evaluation, their common purpose in the context of this research was to assess the extent to which development projects have consistently achieved the intended change for social betterment.

However, whilst there is no universal agreement on one specific definition itself, it is clear from the perspective of evaluation of development interventions that funders are mainly interested in evaluations, which can judge the impact and effectiveness of their funding support. The diverse definitions of evaluation have also culminated in a variety of models of evaluation, which reflect different perspectives of various theorists regarding the purpose and functions of an evaluation.

Owen and Rogers (1999) developed a meta-model consisting of five categories of evaluative enquiry, which are designed to address the "why" (why an evaluation is conducted) question from the perspective of an evaluation practitioner. The five categories of evaluative enquiry, which form the theoretical grounds for conducting good evaluations include: "proactive, clarificative, interactive, monitoring and impact and that each of them produces useful knowledge for decision-making" (Owen and Rogers (1999, p.60).

The 'clarificative' form of evaluative enquiry focuses mainly on explaining the "internal structure and functioning of an intervention, the collection and analysis of data, the intended outcomes and program design, the underlying rationale, and plausibility of the program" (Owen and Rogers 1999, p.42-43). Owen and Rogers (1999) further explain that 'interactive' form of evaluation is based on the assumption that those with direct vested interest in the intervention should also control the evaluation of the intervention; the 'monitoring' form of evaluative enquiry is undertaken when an intervention is ongoing and is intended to assess the progress of an intervention.

Finally, 'impact' evaluation is used to assess the effects of a completed programme and is intended to inform decision-making on the worthiness of an intervention. For purposes of this research, the clarificative form of evaluative enquiry was adopted here as an appropriate evaluation framework for this metaevaluation study because of its focus on issues of programme design, internal structure, objectives, rationale and outcomes, which are in alignment with both the universally accepted PES and the thrust of this metaevaluation research.

Metaevaluation is grounded in the theory for good programme evaluation (Shadish, Cook and Leviton, 1991; Alkin, 2012; Ambron, Dornbusch, Hess, Hornik, Philips, Walker and Weiner, 1980). This theory is based on the vision of improving society's well-being. This vision, through evaluation, is idealised in the context of solving the problems that society faces.

Shadish, Cook and Leviton, (1991) for example, demonstrate that the process of solving the problems of society is guided by a problem-solving sequence. These authors explain that this sequence, involving the identification of a problem in society, generating and implementing alternatives to reduce the adverse effects caused by the problem, evaluating the effectiveness of these alternatives, and adapting alternatives that the evaluation results suggest, can significantly reduce the problem in society.

Furthermore, Alkin (2012) demonstrates that the evolution of the theory for good evaluation was driven mainly by the quest to promote accountability and social enquiry. This evolutionary process subsequently precipitated a surge in theoretical debates, which focused mainly on three major considerations related to evaluation methodology, the manner in which data are to be valued and judged, and the uses of the results of the evaluation process.

Shadish, Cook and Leviton (1991) for example, compound the three major considerations of use, methods and valuing into three stages of theories. According to these authors, first stage theories emphasise the importance of social programming and the discovery of knowledge; second stage theories focus on the way evaluation is used and its social utility; and third stage theories address the integration of inquiry and utility. Grounded in these three stages of theories, Stufflebeam (1974) suggests a range of conceptualisation premises that can be used to conduct metaevaluation activities. Six conceptualisation premises were adopted from Stufflebeam (1974) with modifications for the purposes of this research based on the three stages of theories. These were:

1. Metaevaluation should assess the merit and worth of evaluations using appropriate criteria.
2. Metaevaluation should assess the objectives, designs and quality of an evaluation.
3. Metaevaluation should describe and judge evaluation work and recommend how evaluations can be improved and how the findings can be used appropriately.
4. Metaevaluation should serve evaluators and all the persons who are interested in their work.
5. Metaevaluation must collect, organise and analyse necessary information and apply it to appropriate decision-making and learning processes.
6. Metaevaluations must be technically adequate and useful.

Since the development of the programme evaluations standards in 1994 by the Joint Committee on evaluation standards, there has been an upsurge of interest in the quality of evaluations of social interventions (The Joint Committee, 1994). Metaevaluation is used to assess the quality of either a single evaluation or a set of evaluations. Hanssen, Lawrenz and Dunet (2008) deduce that just as evaluation can help improve programmes and contribute to knowledge in the field, conducting metaevaluation can improve the quality of evaluations, thereby enhancing the practice of evaluation.

Stufflebeam (2001) argues that the field of evaluation has advanced sufficiently in its methodology and public service that evaluators can and should subject their evaluations to systematic metaevaluation. In addition, Pierre (1983) argues that metaevaluation should also be concerned with the re-analysis of data generated in evaluations in order to verify the credibility of early findings, and clarify the inferences that may have been drawn. Fitzpatrick, Sanders and Worten (2004) and Hanssen, Lawrenz and Dunet (2008) all argue for the increased use of metaevaluation. In particular, Fitzpatrick, Sanders and Worten (2004) re-iterate the lack of the use of metaevaluation and argue for its increased use in improving evaluation conduct and practice. Furthermore, Stufflebeam (2001), and Hanssen, Lawrenz and Dunet (2008) strongly recommend the increased use of metaevaluation to describe and judge information concerning the strengths and weaknesses of an evaluation.

Metaevaluation has been utilised in different developed country contexts to assess the quality of evaluations. Usikyla and Virtanen (2000), for example, used metaevaluation to determine the quality of a case study of 15 evaluations of different European Structural Fund Programmes in Finland. Odom and Fewell (1983) conducted a metaevaluation of an evaluation of the Early Childhood Special Education programme in the USA. Apthorpe and Gasper (1982) used metaevaluation to assess the credibility of the evaluation of rural co-operatives in the Netherlands.

While metaevaluation has been utilised in different developed country contexts to assess the quality of evaluations, the application of metaevaluation to assess the quality of evaluations of development interventions in developing countries remained scanty. What, then, could be learnt from a developing country like Uganda? Situated in East Africa, along the equator, Uganda, a British protectorate until its independence on 9th October 1962, was regarded by Sir Winston Churchill as an economic success story and the pearl of Africa (Brett, 1996, Churchill, 2015).

During the time of its independence, Uganda had achieved rapid growth in its agricultural sector and enjoyed a vibrant industrial sector. These achievements, adds Brett (1996), were mainly supported by competent economic management plus a favourable climate and fertile soil at the time of Uganda's independence. Uganda also turned out to be a significant advocate of the World Bank's programmes in Africa. This country also became the first developing country to complete the debt reduction package and was the first African country to have its Poverty Reduction Strategy Paper (PRSP) approved enabling the country to qualify for enhanced debt relief (World Bank 2009). Since 1987, the World Bank has provided the Ugandan government with more than US$790 million in loans. In addition, an estimated US$1 billion grants in project support have also been provided to the Ugandan government to support agriculture, infrastructure, export-led industrial growth and social sectors (World Bank 2009).

In 1994 Uganda was one of nine African countries (the others being Burkina Faso, Gambia, Ghana, Guinea, Lesotho, Nigeria, Tanzania and Zimbabwe) declared as one of Africa's economic success stories in completing adjustment programmes with increasing returns to sustainable growth and poverty-reduction levels (World Bank, 1994). Uganda continues to receive close attention from the World Bank and attracts generous development loans from the latter through its PRSCs towards its poverty reduction efforts.

The PRSCs upon which the selected evaluation cases are based were implemented at different intervals in Uganda. The first case was an evaluation of the first poverty reduction intervention funded through the first Poverty Reduction Support Credit (PRSC1) under Credit Number 3510, which was approved on 31st May 2001, became effective on 29th November 2001 and closed on 31st March 2002 (World Bank, 2009). The second poverty-reduction intervention, funded through the second Poverty Reduction Support Credit (PRSC2) under Credit Number 3696, which continued to support Uganda's Poverty Eradication Action Plan (PEAP) of the year 2000 was approved on 23rd July 2002, became effective on 12th March 2003 and closed on 30th June 2003.

The third poverty-reduction intervention, funded through the third Poverty Reduction Support Credit (PRSC3) under Grant Number H0630 was approved on 9th September 2003, became effective on 11th February 2004 and closed on 30th September 2004. The fourth poverty-reduction intervention, funded through the fourth Poverty Reduction Support Credit (PRSC4) under Grant Number H1220UG was approved on 2nd September 2004, became effective on 28th February 2005 and closed on 30th September 2005.

In summary: the first selected evaluation case was based on the first Poverty Reduction Support Credit (PRSC1) under Credit Number 3510 which was implemented from 29th November 2001 to 31st March 2002; the second selected evaluation case was based on the second Poverty Reduction Support Credit (PRSC2) under Credit Number 3696 which was implemented from 12th March 2003 to 30th June 2003; the third selected evaluation case was based on the third Poverty Reduction Support Credit (PRSC3) under Grant Number H0630 which was implemented from 11th February 2004 to 30th September 2004 and the fourth selected evaluation case was based on the fourth Poverty Reduction Support Credit (PRSC4) under Grant Number H1220-UG which was implemented from 28th February 2005 to 30th September 2005.

The PRSCs were project-specific and the priority sectors, among others, included education and health. The implementation of PRSCs followed a period in which Uganda was the first developing country to complete the debt-reduction package and was also the first to have its Poverty Reduction Strategy approved in the year 2000, enabling the country to qualify for enhanced debt relief during this same year due to its strict adherence to the fiscal discipline requirements and its sound economic reform record.

In general, the evaluation of the effectiveness of poverty-reduction interventions in Uganda has been conducted by various organisations since the completion of the debt-reduction package as well as the approval by the World Bank of Uganda's Poverty Reduction Strategy in 2000. The United Nations Development Programme's (UNDP) Human Development report, for example, showed that in 2005 more than 20% of Ugandans lived below US$1.25 dollar a day and inequality had worsened where the income difference (measured by the Gini Coefficient) worsened from 0.33 to 0.43 (UNDP 2005).In addition, the assessment of poverty and vulnerability by the World Bank in 2006 demonstrated an increase in the poverty rate (World Bank, 2006; Sala-i-Martin, 2006).

Furthermore, the evaluation of the World Bank's overall strategic African development Action Plan (AAP) from 2001 to 2007 adjudged specifically that, although Uganda had posted robust economic growth during the same period, these economic growth rates had not, paradoxically, translated into overall reduction in the incidence of poverty (World Bank 2009).

With specific reference to sector-specific PRSCs in Uganda, evaluations of the World Bank's support to Uganda through the PRSCs on sector-specific interventions to improve, for example, water and sanitation specifically as part of its human development agenda showed satisfactory performance" (World Bank, 2009, p.17). What then could be learnt specifically from selected World Bank evaluation cases, which assessed the effectiveness of the World Bank's poverty-reduction interventions, funded through the PRSCs in Uganda from 2001 to 2005?

With scant evidence in the literature from Uganda on the utilisation of metaevaluation to assess the quality of evaluations of poverty-reduction interventions, it was demonstrated in the preceding sections that metaevaluation has been mainly used in various evaluation case studies in developed countries to determine the quality of evaluations.

This study on metaevaluation of World Bank evaluations of poverty-reduction interventions in a developing country context of Uganda used four sector-specific completed evaluations of the World Bank on sector-specific poverty-reduction interventions funded through its PRSCs. These selected PRSC evaluation cases for this thesis were released to the public by the World Bank in 2009 (World Bank 2009). In determining evaluation outcomes of the effectiveness of development interventions, the World Bank uses three evaluation dimensions to assess the extent to which an intervention has achieved its stated objectives. These evaluation dimensions are relevance, efficacy and efficiency.

Relevance includes the relevance of objectives and of designs. Relevance of objectives is the extent to which the project's objectives are consistent with a country's development priorities (in this case poverty-reduction priorities in accordance with Uganda's poverty-reduction strategy papers), and relevance of design is the extent to which the project's design is consistent with the stated objectives. Efficacy is the extent to which the project's objectives were achieved, and efficiency is the "extent to which the project achieved a return higher than the opportunity cost of benefits at the least costs compared to alternatives" (World Bank 2009, p.23). The scrutiny of evaluation outcomes revealed evidence of remarkable inconsistencies in evaluation outcomes in the effectiveness of all four selected poverty-reduction interventions of the World Bank in Uganda funded through its PRSCs during the period 2001 to 2005. Evaluation outcomes across the four selected evaluation cases funded through its PRSCs provided glaring evidence of inconsistencies in evaluation outcomes in the effectiveness-selected poverty-reduction interventions during the period from 2001 to 2005.

Moreover, in practice development funders are mainly interested in steady consistency in the effectiveness of their funding support. However, what was particularly clear from these evaluation outcomes was that there was evidence of inconsistent outcomes in the effectiveness of the selected poverty-reduction interventions funded by the World Bank through its poverty reduction support credits (PRSCs).The evidence from the assessment of evaluation outcomes showed that expected impacts of crossing-cutting reforms did not materialise as, in some cases, the reforms, for example, were not carried out in the case of the first Poverty Reduction Support Credits (PRSC1).

In terms of rural development outcomes, the evidence from evaluation outcomes revealed that the increase in the investment budget for agriculture, which was deemed as a necessary and important step for generating and disseminating technology, was insufficient to change living conditions in the short-term and little changed in terms of transparency and reduced corruption in the case of the second Poverty Reduction Support Credits (PRSC2).

In the cases of PRSC3 and PRSC4, evaluation outcomes in the effectiveness of poverty-reduction interventions revealed that the interventions did not lead to more efficient and equitable use of public resources nor did they to the cross-cutting reforms that could improve service delivery. These evaluation outcomes in the effectiveness of poverty-reduction intervention in the cases of PRSC1, PRSC2, PRSC3 and PRSC4, as discussed in detail in Chapter Two on literature review, were inconsistent with the satisfactory evaluation results of the World Bank's poverty reduction interventions funded through these PRSCs in specific sectors, such as health, economic growth, public reform, agriculture, education, and water and sanitation based on the four selected evaluation cases in Uganda.

The dynamic assortment of evident inconsistencies in evaluation outcomes in the effectiveness of selected evaluation cases of poverty-reduction interventions in Uganda, as well as the demand by development funders for steady consistency in the effectiveness of their funding support, raised serious questions about the legitimacy and quality of these selected evaluations and the evidence used to support decision-making. The questions about the legitimacy and quality of these selected evaluations implicitly heightened the importance of evaluating the quality of these evaluations of poverty-reduction interventions funded by the World Bank through its PRSCs.

The evaluation literature promotes the importance of evaluating evaluations in order to determine the legitimacy and quality of evaluations. Cooksy and Caracelli (2005) claimed that evaluation results influence decision-making processes that are aimed at improving the effectiveness of social interventions, and that the capability for continuous improvement of social interventions can be undermined when evaluations are flawed.

In order to improve the effectiveness of social interventions consistently, the concern with evaluating the quality of evaluations of social interventions is of paramount importance. The evident inconsistencies in the evaluation outcomes of poverty reduction interventions funded by the World Bank through its PRSCs, coupled with scanty evidence in the literature on the utilisation of metaevaluation to assess the quality of evaluation of development interventions in Uganda, which identified and resolved the causes of these inconsistencies, raised serious questions about the overall quality of evaluations and the evidence, which was used to support judgments on the effectiveness of these poverty-reduction interventions in Uganda. What then could be learnt from the metaevaluation of evaluations of specifically poverty-reduction interventions funded by the World Bank through the PRSCs in Uganda?

1.1.2 Problem Statement, Knowledge Gap and Purpose

It was evident from the preceding paragraphs that poverty-reduction interventions funded through the PRSCs by the World Bank in Uganda had been systematically evaluated in order to measure their effectiveness. However, overall evaluation evidence on the effectiveness of these poverty interventions in Uganda suggested remarkable inconsistencies in the evaluation outcomes of the effectiveness of these poverty interventions.

Moreover, scanty evidence existed in the literature on Uganda, which evaluated the quality of evaluations of poverty-reduction interventions funded through the PRSCs by the World Bank in order to identify and resolve the causes of these inconsistencies in the evaluation outcomes of the effectiveness of such poverty-reduction interventions.

The evident inconsistencies in the evaluation outcomes, coupled with scanty evidence on substantial attempts to identify and resolve the causes of these inconsistencies, raised serious questions about the overall legitimacy and justifiability of evaluation evidence, which was used to support judgments on the effectiveness of poverty reduction interventions of the World Bank in Uganda.

The purpose of this research was, therefore, to evaluate the quality of selected evaluations of the World Bank funded through its PRSCs in Uganda so that evident causes of such remarkable inconsistencies in evaluation outcomes could be clarified and resolved.

1.1.3 The Research question

In order to examine the quality of selected evaluations of poverty-reduction interventions, which were funded by the World Bank through its PRSCs in Uganda so that evident causes of such remarkable inconsistencies in evaluation outcomes could be identified and resolved, the following research question was specified:

1) How effectively did selected evaluations conducted under the auspices of the World Bank on poverty-reduction interventions funded through the PRSCs in Uganda meet the requirements of a good evaluation?

In the recent years, there has been an upsurge, since the publication of the PES in 1994, of interest in the quality of evaluations in order to enhance the effectiveness of development interventions (Bamberger, 2006; Stufflebeam, 2001; Schwandt and Halpern, 1988; Stufflebeam, 1981; Stufflebeam, 1978). Stufflebeam (2001), for example, argues that the “evaluation field has advanced sufficiently in its methodology and public service that evaluators can and should subject their evaluations to systematic metaevaluation” (p. 1).

Therefore, the evaluation of the extent to which primary evaluations meet the requirements of a good evaluation systematically reinforces the importance of subjecting primary evaluations to advanced evaluative scrutiny in order to determine their credibility and legitimacy. The evaluation of how effectively selected evaluations met the requirements of a good evaluation in the case of Uganda was not only intended to assess for consistency in the evaluation outcomes of the effectiveness of the intervention, but also to inspire future use of metaevaluation as a means for identifying and resolving evident causes of inconsistencies in the effectiveness of development interventions in other developing country contexts.

The evaluation of the quality of evaluations sought to provide greater clarity on the underlying causes of inconsistencies in evaluation outcomes and enhance the performance of development interventions, inform policy and the kinds of policy change. This metaevaluation exercise also sought to reveal evidence, which could have had the potential of influencing thinking about the possibilities of holistic change in the programmatic architecture of the intervention.

1.1.4 Research Methodology

The methodology for evaluating the quality of four selected evaluations of poverty reduction interventions funded by the World Bank through its PRSCs was based on the mixed methods approach, grounded in constructivist evaluation within the context of the constructivist paradigm. The methodology was composed of case studies, document analysis and descriptive frequency statistics.

The selected methods assisted in identifying evident causes of such remarkable inconsistencies in evaluation outcomes of selected evaluation cases and answer the specified research question. Selected mixed methods provided the appropriate leverage for triangulating multiple perspectives and information sources in order to generate credible evidence, identify evident causes of such remarkable inconsistencies and answer the specified research question. Triangulation also insulated the study from the generation of evidence based on what could have been construed as pieces of evidence based on a single method. The use of mixed methods also protected the evaluation process of the quality of evaluations from probable bias associated with the use of a single method to generate empirical evidence.

Drawing from an assortment of information sources based on selected evaluation cases, a case study approach was appropriate in addressing information validity imperatives because of attendant multiple perspectives evident from such multiple cases. The closeness of selected evaluation cases to real–life situations provided a wealth of details and useful strands of knowledge. In addition, the use of multiple case studies improved the credibility of evidence and insulated the generation of evidence from what would be construed as biased explanations based on one single method. The use of document analysis involved identifying, choosing, evaluating and integrating data obtained from evaluation reports of the four selected evaluation cases.

The prescribed quantitative scales based on the quality checklist prescribed under the PES were used for scoring the quality of evaluations. Determinations regarding the extent to which each selected case complied with the prerequisites of a credible evaluation stipulated under the PES were made based on a set of prescribed quantitative scales and scores. Each of the four selected evaluation cases in Uganda's case study was subjected to a checklist of 300 requirements per evaluation case and was composed of 1,200 checklists for all four selected evaluation cases. Four compulsory associated standards specified under the PES were: Service Orientation (P1), Valid Information (A5), Justified Conclusions (A10) or Impartial Reporting (A11) on any of the four compulsory associated standards.

On the basis of document analysis, determinations regarding the extent to which each selected case complied with the prerequisites of a credible evaluation under the PES were made based on the following quantitative scales: 0-2: Poor, 3-4: Fair, 5-6: Good, 7-8: Very Good and 9-10: Excellent (Stufflebeam, 1999, p.2; see also Appendix C on page 202 and Appendix D on page 205, respectively).

Descriptive frequency statistics were also used to summarise the data, which was generated from document analysis using Statistical Package for Social Sciences (SPSS) based on the above prescribed quantitative scores stipulated under the PES. Summarised emerging descriptive frequency patterns supported the metaevaluation process in generating new insights into the behaviour of outcomes and assisted in creating meaningful interpretations of the results of the assessment.

1.2 Chapter Two: Literature Review.

Metaevaluation was situated for purposes of this study in the clarificative form of evaluative inquiry as an appropriate metaevaluation framework because the clarificative form of evaluative inquiry focuses on issues of program design, internal structure, objectives, rationale, and outcomes which are in alignment with both the universally accepted program evaluation standards (PES) and the main thrust of this research.

Metaevaluation for purposes of this study was also grounded in the theory for good programme evaluation based on the vision of improving the wellbeing of society, and idealized in the context of solving the problems that society faces which, involves the identification of a problem in society, generating and implementing alternatives to reduce the adverse effects caused by the problem, evaluating the effectiveness of these alternatives, and adapting alternatives that evaluation results suggest, can significantly reduce the problem in society.

Metaevaluation has been utilised in different developed country contexts, such as The Netherlands and Finland, to assess the quality of evaluations of social development interventions. However, for purposes of this study, it was discovered that there was scant evidence in the use of metaevaluation to assess the quality of evaluations of poverty-reduction interventions in Uganda.

The nature of Uganda's governance discourse was characterised by the omnipresence of the State in every sector of the economy after independence in 1962. The rationale for the intervention of the World Bank in Uganda's development processes mainly aimed at transforming the State and the economy which focused primarily on trimming bloated State bureaucracies to spur sustainable economic and social development through the implementation of various development programmes, such as poverty-reduction interventions which were funded through the World Bank's PRSCs which were project-specific and the priority sectors, among others, included education, public service reform and health.

The evaluation outcomes of the effectiveness of the four selected poverty-reduction interventions of the World Bank funded through PRSCs were scrutinised for purposes of this research. Evidence of an assortment of remarkable inconsistencies in evaluation outcomes in the effectiveness of selected poverty-reduction interventions was uncovered and yet, in practice, development funders demand steady consistency in the effectiveness of their funding support.

The dynamic assortment of evident inconsistencies in evaluation outcomes in the effectiveness of selected evaluation cases of poverty-reduction interventions in Uganda, as well as the inevitable demand by development funders for steady consistency in the effectiveness of their funding support, raised serious questions about the legitimacy and quality of these selected evaluations, as well as the evidence used to support decision-making. Questions about the legitimacy and the quality of these selected evaluations implicitly heightened the importance of evaluating the quality of these evaluations of poverty-reduction interventions.

1.3 Chapter Three: Research Methodology

The research methodology was based on a mixed methods approach composed of case studies, document analysis and descriptive frequency statistics grounded in constructivist evaluation within the context of the constructivist paradigm using the evaluation quality checklist prescribed by the PES. The use of mixed methods allowed for triangulating a multiplicity of these selected methods, which provided multiple sources of reliable evidence, enhanced the credibility of evidence and assisted in identifying evident causes of such remarkable inconsistencies in evaluation outcomes of selected evaluation cases in order to provide plausible answers to the research question.

The use of document analysis involved identifying, choosing, valuing and integrating data obtained from documents using the four selected evaluation cases. The prescribed quantitative scales based on the evaluation quality checklist prescribed under the PES were used for scoring the quality of evaluations. The judgments regarding the adequacy and strength of each selected evaluation case in meeting the requirements of a good evaluation based on the PES were made based on a set of prescribed quantitative scores.

Descriptive frequency statistics were also used to summarise the data, which were generated from document analysis using computerised SPSS based on the above prescribed quantitative scores generated from stipulated quantitative scales under the PES. Summarised emerging frequency patterns from prescribed quantitative scores supported the evaluation process in generating new insights into the behaviour of outcomes and assisted me in creating meaningful interpretations of the results of the assessment based on emerging patterns.

1.4 Chapter Four: Presentation of Findings

General findings from document analysis without specific reference to specified compulsory associated standards, based on the PES, showed that the strengths of the first and fourth evaluation cases were found to be poor on the basis of the Stakeholder Identification (US1) criterion. The findings from document analysis also revealed that persons who conducted the evaluation were not fully competent to perform the evaluation according to the Evaluator Credibility (US2) criterion. In terms of Defensible Information Sources (AS4) criteria, the findings revealed that the strengths of both the first and four evaluation cases were also found to be poor.

Evidence from the study showed that the strengths of the second and third evaluation cases were poor on the reliable information (AS6) measure. The findings demonstrated that quantitative information was not appropriately and systematically analysed so that evaluation questions could be answered effectively, and on the basis of the analysis of the quantitative information (AS8) measure, it was demonstrated in the study that the strengths of the first and third evaluation cases were also poor.

Finally, in terms of the Metaevaluation (AS12) criterion, which stipulates that an evaluation itself should be formatively and summatively evaluated against these PES and other pertinent standards, so that its conduct is appropriately guided, and on completion, stakeholders could closely examine its strengths and weaknesses. The findings from document analysis demonstrated that the strengths of all the four selected evaluation cases were found to be poor on the Metaevaluation (AS12) criterion.

With reference to specified compulsory standards under the PES, it was also revealed in the thesis that notwithstanding significant variations in metaevaluation findings, the strength of each of the four selected evaluation cases was not found to be poor, and, therefore none of the selected evaluation cases could be rejected based on the compulsory evaluations standards specified under the PES. In terms of summarised descriptive frequency statistics, the findings revealed that Report Clarity (US5) was more important across the four selected evaluation cases than the need to identify a full range of stakeholders [Stakeholder Identification (US1)] and that Practical Procedures (FS1) were more important than political viability [(Political viability (FS2)] across all four selected evaluation cases.

Summarised descriptive frequency statistics also showed that formal descriptions of sources of information [Programme documentation (AS1)] were more important across all the four evaluation cases than obtaining information from defensible sources [Defensible Information Sources (AS4)]. Summarised descriptive frequency statistics also showed that the acquisition of systematic information [Systematic Information (AS7)] was more important than ensuring the quality of evaluations [Metaevaluation (AS12)] across all four selected evaluation cases.

Summaries of overall findings in general from document analysis across all four selected evaluation cases are provided in Table F1 (p.225).The findings from document analysis across all four selected evaluation cases on compulsory PES, which will be analysed in detail in Chapter Five, are presented in Table G1 (p.227). Summaries of findings based on descriptive statistics across all four selected evaluation cases, as well as all four categories of evaluation standards under the PES, which will also be analysed in detail in Chapter Five, are provided in Tables H1, H2, H3 and H4 (p.228).

1.4 Chapter Five: Analysis of the Findings and Answers to the Research Question

1.4.1 Analysis of the Findings: The analysis of the findings showed that a range of persons involved in, or affected by the evaluations under consideration were not fully identified and their needs were not adequately addressed. The analysis of findings from document analysis showed that appropriate procedures of gathering information were not well selected and implemented to ensure that the information, which was obtained during the evaluation process, was sufficiently reliable for the intended use, and that quantitative information was not appropriately and systematically analysed so that evaluation questions could be answered effectively. The analysis revealed that selected evaluation themes were not "formatively and summatively evaluated against pertinent standards so that their conduct could be guided appropriately and on completion, stakeholders could closely examine their strengths and weaknesses" (Sanders, 1994, p.23).

1.4.2 Answers to The Research Question: How effectively did selected evaluations conducted under the auspices of the World Bank on poverty-reduction interventions projects funded through the PRSCs in Uganda meet the requirements of a good evaluation based on the PES? The analysis provided significant lessons, which demonstrated that there were evident variations in the extent to which each of the four selected evaluation cases complied with the prescribed requirements of a good evaluation based on the PES. On overall, the analysis of the findings revealed that the four selected evaluations cases performed better on bureaucratic criteria and worse on accountability and transparency procedures.

In particular, the evidence from the analysis surprisingly revealed that a full range of stakeholders was not adequately identified and engaged in both designing and the evaluation processes of selected interventions, which perilously exacerbated remarkable inconsistencies in the evaluation outcomes of the effectiveness of poverty reduction interventions and undermined the quest to improve accountability, transparency and social enquiry thus constraining the ability to resolve the problem of bureaucratic inefficiency and thus undermining the use of substantive participatory imperatives.

Notwithstanding these variations in the extent to which each of the four selected evaluation cases complied with the requirements of a good evaluation based on the PES, the analysis revealed that the strength and quality of each of the four selected evaluation cases were not found to be poor based on compulsory associated standards of a good evaluation prescribed under the PES.

1.5 Chapter six: Conclusions, Contribution to Knowledge and Areas for further Research.

1.5.1. Conclusions: The evidence from the study, on overall, revealed that the four selected evaluations cases performed better on bureaucratic criteria and worse on accountability and transparency procedures. In particular, the evidence from the analysis surprisingly revealed that a full range of stakeholders was not adequately identified and engaged in both designing as well as in the evaluation process of selected interventions, which is inconsistent with the essential principles of accountability, transparency and participatory social enquiry undergirding the theory for good programme evaluation upon which this research was grounded.

The absence of the engagement of a full range of stakeholders undoubtedly exacerbated remarkable inconsistencies in the evaluation outcomes of the effectiveness of poverty-reduction interventions. The identification and engagement of a full range of stakeholders enhance broad decision-making, thus, improving accountability and transparency, and participatory social enquiry and resolves the problem of bureaucratic inefficiency.

1.5.2 Contribution to Knowledge: The researcher makes one contribution to metaevaluation methodology and submits that whilst stakeholder identification which is grounded according to Owen and Rogers (1999) in the interactive style of evaluative inquiry based on the assumption that those with direct vested interests in the intervention should also control the evaluation of the intervention is not prescribed as one of the compulsory associated standards based on the PES developed in 1994 by the Joint Committee on Evaluation Standards to measure the quality of evaluations, it is empirically plausible to elevate stakeholder identification as an additional independent compulsory associated standard under Utility Standards based on the PES when conducting metaevaluations of not only poverty-reduction interventions, but also of other social development interventions.

1.5.3: Areas for further Research: The researcher proposes the following significant problematic areas, which emerged from the study for future research in metaevaluation, which the researcher could not continue undertaking due to logistical and conceptualisation limitations prescribed for this study:

1.5.3.1: Assessment of the Credibility of Evaluators: A metaevaluation of the competencies of persons conducting the evaluation and the identification of the nature of deficiencies in the level of professional trust in persons conducting the evaluation which may have potentially undermined the credibility of evaluation outcomes may contribute towards resolving possible causes in the effectiveness of evaluation outcomes and can also contribute to the improvement of the practice of the evaluation profession.

1.5.3.2: Assessment of Defensible Information Sources: An assessment of the reliability of information sources and the effectiveness with which such sources of information are described when conducting primary evaluations would be central in providing suitable information on the extent to which the evaluation process was supported with defensible sources of information.

1.5.3.3: Assessment of Reliable Information: An investigation into the use of appropriate procedures of gathering information during the evaluation process and the extent to which such information was sufficiently reliable for the intended user of the primary evaluation would contribute towards providing defensible grounds to inform decision-making processes.

1.5.3.4: Analysis of Quantitative Information: An examination of the degree of vigor with which appropriate and systematic quantitative information was analysed when conducting primary evaluations would contribute towards assisting in understanding the extent to which evaluation questions were answered effectively

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

The introduction to the thesis in the summarised version of this research in chapter one provided an overview of the case for metaevaluation within the context of evaluative enquiry and specified the research problem, research gap and purpose in relation to metaevaluation theory, as well as the methodological approach for data collection and analysis.

The overall objective of this chapter is to make a case for metaevaluation and determine the effects of the interventions of the World Bank in reducing poverty in Uganda, so that the basis for the specification and application of the methodology for data collection and analysis for conducting metaevaluation of selected World Bank evaluation cases (reports) on poverty-reduction interventions in Uganda is specified in Chapter Three.

The chapter is divided into three major sections. The first section discusses the importance of metaevaluation within the context of evaluative enquiry. The second section reviews the literature in Uganda's case study; discusses the rationale for the intervention of the World Bank in Uganda; specifies selected evaluation cases; and scrutinises evaluation outcomes of the effectiveness of selected poverty-reduction interventions in Uganda. The conclusions of the chapter are provided in the third section.

2.1 The Case for Metaevaluation

Defined by Scriven (1969) as the evaluation of evaluations, metaevaluation is a unique and specialised form of evaluation, which helps in judging the strengths and weaknesses of a primary evaluation so that interested parties are able to judge the reliability and credibility of evaluations, and the evidence used to make decisions (Stufflebeam, 2000; Stufflebeam, 2001; Stufflebeam and Shinkfield, 2007; Cooksy and Caracelli, 2009).

Since metaevaluation is a form of evaluation within the context of evaluative enquiry, there is a variety of definitions of the term 'evaluation' reflecting different perspectives of various theorists regarding the purpose and functions of an evaluation (Tyler, 1942, Tyler, 1950; Cronbach, 1963; Alkin, 1969; Alkin, 1970; Scriven, 1969; Glass, 1969; Wittrock and Wiley 1970; Stufflebeam, 1971; Nevo, 1983; Borich, 1990). For example, whilst Tyler (1942) propagates the perspective that evaluation is a process of determining the extent to which the objectives of an intervention are realised, Cronbach (1963), Alkin (1969) and Stufflebeam (1971) assert that evaluation is a process for providing evidence for decision-making.

In addition, the Organization for Economic Cooperation and Development (OECD) defines evaluation as the process of determining the worth or significance of an activity, policy or programme (OECD 2004). Moreover, evaluation is defined by the World Bank as "the process of assessing the performance of policies, programmes, projects and processes by identifying and disseminating the lessons learnt from experience and by making recommendations drawn from evaluation findings" (World Bank 2010, p.26).

The diverse definitions of the purpose of evaluation demonstrate that the field of evaluation continues to attract different intellectual orientations. Whilst there are many different theoretical perspectives regarding the purpose of evaluation, the common purpose of these is to assess the extent to which development projects have consistently achieved the intended change for social betterment.

However, whilst there is no universal agreement on any one specific definition itself, it is clear, particularly from the perspective of evaluation of development interventions, that funders are mainly interested in evaluations, which can progressively judge the steady impact and effectiveness of their funding support. The diverse definitions of evaluation have also culminated in a variety of models of evaluation, which reflect different perspectives of various theorists regarding the purpose and functions of an evaluation. Owen and Rogers (1999) explain that a model is a prescription for undertaking an evaluation based on certain theoretical assumptions. Stufflebeam and Webster (1983) clarify that the proliferation in the evaluation models is a result of a range of attempts by evaluation theorists to classify evaluation models in terms of elements, such as assumptions, methodology and the engagement of stakeholders.

However, Orr, Cameron and Day (1991) noted that, although there has been a proliferation in evaluation models, a theoretical framework, which would make sense of this situation from the point of view of the evaluation practitioner, was missing. In order to address this missing link, Owen and Rogers (1999) developed a meta-model consisting of five categories of evaluative enquiry, which are designed to address the "why" (why an evaluation is conducted) question in evaluation from the perspective of an evaluation practitioner.

Owen and Rogers (1999) argue that the five forms of evaluative enquiry demonstrate "five dominant styles of evaluation and that each of them produces useful knowledge for decision-making" (p. 60).The five categories of evaluative enquiry, which form the theoretical grounds for conducting good evaluations include: proactive, clarificative, interactive, and monitoring and impact, and all are briefly discussed here. Owen and Rogers (1999) explain that the 'proactive' form of evaluative inquiry "takes place before a programme is designed and that it is intended to provide input to decisions regarding the best way to develop a programme in advance of the planning stage" (p. 60).

The 'clarificative' form of evaluative enquiry focuses mainly on explaining the "internal structure and functioning of an intervention, the collection and analysis of data, the intended outcomes and program design, the underlying rationale, and plausibility of the program" (Owen and Rogers 1999, p.42-43). These authors further explain that 'interactive' evaluation is based on the assumption that those with direct vested interest in the intervention should also control the evaluation of the intervention; the 'monitoring' form of evaluative enquiry is undertaken when an intervention is ongoing and is intended to assess its progress. Finally, 'impact' evaluation is used to assess the effects of a completed programme and is intended to inform decision-making on the worth of an intervention.

The clarificative style of evaluative enquiry was adopted here as an appropriate evaluation framework for this metaevaluation study because of its focus on issues of programme design, internal structure, objectives, rationale and outcomes, which are in alignment with both the universally-accepted PES and the thrust of this research.

Owen and Rogers (1999) specify that, compared to other forms of evaluative enquiry, such as the proactive form of evaluative enquiry, clarificative evaluation involves the collection and analysis of data, and seeks to address issues, such as whether or not the intended outcomes can be achieved by the nature of programme design, the underlying rationale for the programme, and its plausibility.

2.1.1 The Theoretical Basis for Metaevaluation

Metaevaluation is grounded in the theory for good programme evaluation (Weiss, 1995; Shadish, Cook and Leviton, 1991; Alkin, 2012; Ambron, Dornbusch, Hess, Hornik, Philips, Walker and Weiner, 1980). The theory for good programme evaluation is based on the vision of improving the well-being of society. The vision of improving the well-being of society through evaluation within the context of development is idealised in the context of solving the problems faced by society.

Shadish, Cook and Leviton (1991), for example, demonstrate that the process of solving the problems of society is guided by a problem-solving sequence. These authors explain that the problem-solving sequence involves the identification of a problem in society, generating and implementing alternatives to reduce the adverse effects caused by the problem, evaluating the effectiveness of these alternatives, and adapting alternatives that evaluation results suggest, can significantly reduce the problem in society.

Furthermore, Alkin (2012) demonstrates that the evolution of the theory for good evaluation was driven mainly by the quest to promote accountability and social enquiry. This evolutionary process subsequently precipitated a surge in theoretical debate, which focused mainly on three major considerations related to evaluation methodology, the manner in which data are to be valued and judged, and the uses of the results of the evaluation process.

Shadish, Cook and Leviton (1991), for example, compound the three major considerations of use, methods and valuing into three stages of theories. According to these authors, first stage theories emphasise the importance of social programming and the discovery of knowledge; second stage theories focus on the way evaluation is used and its social utility, and third stage theories address the integration of enquiry and utility.

In the context of these three major theoretical considerations, a good theory of programme evaluation emerged, which suggests feasible practices that can be used to construct knowledge of the value of social programmes, which in turn can be used to ameliorate the social problems to which programmes are relevant. The emergent social programming component of the theory assumes that social problem-solving can be improved by incremental improvements in existing programmes, better designing of new programmes or terminating bad programmes and replacing them with better ones (Shadish, Cook and Leviton,1991).

The knowledge component of a good theory of social programme evaluation addresses the contestation between common sense and systematic observations, and, provides answers regarding appropriate methods of constructing knowledge by answering questions, such as: “Why employ professional evaluators? Why not rely on gossip, cronyism, or newspaper reports or lobbyists? Do evaluators offer different or better knowledge than what is already available”? (Shadish, Cook and Leviton, 1991, p.42; Alkin; 2012).

Shadish, Cook and Leviton (1991) and Alkin (2012) advance the evaluative narrative which propagates the understanding that the process of evaluation is a human investigative endeavour, which involves evaluative complexities that are interdependent and often characterised by nonlinear set of problem solving activities. Therefore, the three major considerations of use, methods and valuing demonstrate that humans identify options for improving their own lives by detecting the change, generating pieces of information about the effectiveness of the evaluand (for example, a project, programme, process or organization), which can be used to reduce societal problems within a given context throughout the evaluation process.

In addition, there are also debates within the context of different forms of human behaviours as to whether or not evaluations should be values-free. While Scriven (1966) suggests that evaluations should be value-free in constructing knowledge, Shadish, Cook and Leviton (1991), further, argue that social programmes are not, in themselves, values free, and that it is, therefore, impossible to evaluate in the political world of social programming without taking into account the importance of societal values.

In addition, since practitioners are action-oriented, they orientate their evaluation tasks with pragmatic concepts and methods, which serve as a useful guide in the execution of the tasks (Shadish, Cook and Leviton, 1991). These authors further propose that in order to avoid undertaking evaluations that are incomplete, impractical or technically inferior, the discussion of the good theory of evaluation is important for evaluation practitioners. This metaevaluation research adopted this theoretical basis for good programme evaluation driven mainly by the quest to promote accountability and social enquiry as articulated by Alkin (2012) and compounded by Shadish, Cook and Leviton (1991).

The researcher argues that such human investigative endeavour is grounded in the three major considerations of use, methods and valuing mentioned above and that, the process of evaluation often results in pieces of evidence which inherently reflect a variability of constructions, judgments and envisaged outcomes based on differing forms of behavior.

The reflections of mental constructions and judgements on the effectiveness of the evaluand resulting from such forms of differing behavior are in compliance with the constructivist philosophical paradigm discussed in detail under section 3.1 of Chapter Three which advances the argument that humans generate knowledge and meaning based on their own interactions, experiences and their ideas. For example, the use component of the good theory of evaluation implies that the evaluation results are used to contribute towards solving societal problems.

In addition, underlying the use component of the good theory of evaluation, by extrapolation, implies that social interventions that are not effective in contributing to social betterment based on evaluation results can be terminated and replaced with better interventions. The practice component of the good theory of evaluation helps human investigative processes in ranking the evaluands which have the greatest and most effective opportunity to improve the well-being of society.

2.1.2 Conceptualisation of Metaevaluation

Scriven (1969) first coined the term "metaevaluation" as a concept in the late 1960s. The term 'evaluation of evaluations', as defined by Scriven (1969), was created by Orata in 1940 (Hedler and Gibram, 2009). It is important to note from the outset that metaevaluation does not mean meta-analysis. Weiss (1998) defines meta-analysis as the "systematic summary of the results from a number of different evaluations of the same kind of programs" (p. 48).

The systematic summary of the results from a number of different evaluations of the same kind, involves collecting and summarising empirical data from similar evaluation studies. This process uses statistical approaches to combine the results from multiple evaluation studies of the same kind of programmes in an effort to increase statistical power in order to resolve improbabilities in circumstances where evaluation evidence from similar multiple evaluation studies differ. Ross and Freeman (1999), further explain that meta-analysis tries to determine clusters of findings in order to generate knowledge. Stufflebeam (2001), however, points out that “although metaevaluation and meta-analysis refer to different concepts, they are often inappropriately equated” (p. 187). In clarifying the key difference between the two concepts, Stufflebeam (2001) describes metaevaluation as an assessment of the merit and worth of a given evaluation, whereas meta-analysis is explained as a form of quantitative synthesis of studies that address a similar research question.

Cooksy and Caracelli (2009) describe metaevaluation as a systematic review of evaluations, so that the quality of evaluations and the effectiveness of evaluation processes can be determined. However, while Cooksy and Caracelli (2009) describe metaevaluation as a systematic review of evaluations, it can be added here too that evaluation systems and the use and effectiveness of evaluation tools can also be subjected to metaevaluation in order to, strengthen planning and management of evaluations in organisations. The process of metaevaluation can be carried out formatively when a primary evaluation is being conducted and summatively after a primary evaluation has been completed on effectiveness of an intervention.

In addition, Stufflebeam (2000) explains that metaevaluation involves the “process of delineating, obtaining, and applying judgmental information about the utility, feasibility, propriety and accuracy of an evaluation to report on its strengths and weaknesses” (p.95). Stufflebeam (1974) suggests a range of conceptualisation premises that can be used to conduct metaevaluation activities. Six conceptualisation premises are adopted from Stufflebeam (1974), with modifications, for the purpose of this research. These conceptualisation premises are:

1. Metaevaluation should assess the merit and worth of evaluations using appropriate criteria.

2. Metaevaluation should assess the objectives, designs and quality of an evaluation.

3. Metaevaluation should describe and judge evaluation work and recommend how evaluations can be improved and how the findings can appropriately be used.

4. Metaevaluation should serve evaluators and all the persons who are interested in their work.

5. Metaevaluation must collect, organise and analyse needed information and apply it to appropriate decision-making and learning processes.

6. Metaevaluations must be technically adequate and useful.

In summary, while these conceptualisation premises served as technical hinges around which this metaevaluation study was organised, the researcher points out that the effectiveness of the outcomes of this metaevaluation study depended largely on the extent and the effectiveness with which empirical evidence was generated, triangulated and used to provide multiple sources of reliable evidence. This guided the researcher in constructing plausible knowledge cues and conclusions about the overall outcomes of the study based on four selected evaluation cases, document analysis and descriptive frequency statistics discussed in detail in the next Chapter on research methodology.

2.1.3 Importance of Metaevaluation

Since the development of the programme evaluations standards by the Joint Committee on PES in 1994 (The Joint Committee, 1994), there has been an upsurge of interest in the quality of evaluations of social interventions. Metaevaluation is used to assess the quality of either a single evaluation or a set of evaluations. The literature on metaevaluation outlines two broad categories of metaevaluation, which can be used to assess the quality of a single evaluation or a set of evaluations.

Stufflebeam and Shinkfield (2007) label these two categories as ‘formative’ and ‘summative’ metaevaluations. These authors explain that formative evaluations are intended to improve decision-making through helping evaluators to improve their planning and data collection approaches during the process of evaluating. In addition to improving decision-making, it can be added that summative metaevaluation is also used to judge the strengths and weaknesses of a primary evaluation, so that readers can judge the reliability and credibility of primary evaluations.

Fitzpatrick, Sanders and Worten (2004), and Stufflebeam and Shinkfield (2007) explain that formative metaevaluation provides guidance on the quality of programme evaluations whilst in the process of evaluation. These authors argue that summative metaevaluation helps in preventing the dissemination of invalid conclusions, increasing the utilisation of the results of the primary evaluation and improving the cost-effectiveness of the evaluation. Summative metaevaluations are conducted following an evaluation (Stufflebeam and Shinkfield, 2007).

While Fitzpatrick, Sanders and Worten (2004), and Stufflebeam and Shinkfield (2007) explain that summative metaevaluation prevents the dissemination of invalid results of the primary evaluation, the researcher in addition argues that summative metaevaluation, overall, is used to validate primary evaluations, add credibility to them and enhance the confidence of the users of the results of these primary evaluations. In addition, summative metaevaluations help decision-makers to determine whether or not to make adjustments to the nature of interventions.

Wingate, Gullickson and Cooksy (2010) explain that when a summative metaevaluation finds serious flaws in a primary evaluation, it can prevent decision-makers from taking action based on faulty information. In addition, summative metaevaluation assesses the quality of a completed evaluation, the appropriateness of the evaluation processes and validity of its conclusions, explain (Stufflebeam and Shinkfield, 2007). Furthermore, Greene (1992); and Cooksy and Caracelli (2005, 2009) propagate the understanding that summative metaevaluations helps audiences to see the strengths and weaknesses of a primary evaluation, and to judge its value.

Hanssen, Lawrenz and Dunet (2008) also deduce that, while evaluation helps in proving programmes and as well as learning to enhance the effectiveness of the evaluation profession, conducting metaevaluation improves the quality of evaluations, thus, enhancing the evaluation practice. Stufflebeam (2001) further argues that the field of evaluation has advanced sufficiently in its methodology and public service such that evaluators can and should subject their evaluations to systematic Metaevaluation. In addition, Pierre (1983) argues that metaevaluation should also be concerned with the re-analysis of data generated in evaluations, in order to verify the credibility of early findings and clarify the inferences that may have been drawn.

Fitzpatrick, Sanders and Worten (2004) and Hanssen, Lawrenz and Dunet (2008) all argue for the increased use of Metaevaluation. In particular, Fitzpatrick, Sanders and Worten (2004) reiterate the lack of the use of metaevaluation and argue for its increased use in improving evaluation conduct and practice. Furthermore, Stufflebeam (2001), and Hanssen, Lawrenz and Dunet (2008) strongly recommend the increased use of metaevaluation to describe and judge information concerning the strengths and weaknesses of an evaluation.

Wingate, Gullickson and Cooksy (2010); Stufflebeam and Shinkfield (2007); Greene (1992) and Cooksy and Caracelli (2005, 2009), generally agree in principle that, summative metaevaluation determines the credibility of primary evaluations, improves learning and enhances decision-making, the researcher adapts the general narrative that metaevaluation enhances the effectiveness of both the evaluation practice and evaluation profession. Overall, it is within the context of advancing the evaluation practice and profession that summative metaevaluation was adopted as being the most suitable form of metaevaluation for the purposes of this metaevaluation study.

In practice, metaevaluation has mainly been utilised in different developed country contexts to assess the quality of evaluations. Usikyla and Virtanen (2000), for example, used metaevaluation to determine the quality of a case study of 15 evaluations of different European Structural Fund Programmes in Finland. Odom (1983) conducted a metaevaluation of an evaluation of the Early Childhood Special Education programme in the USA. Apthorpe and Gasper (1982) used metaevaluation to assess the credibility of the evaluation of rural co-operatives in The Netherlands.

Whilst metaevaluation has been utilised in different developed country contexts to assess the quality of evaluations, the application of metaevaluation to assess the quality of evaluations in a developing country context remains scanty. What then could be learnt from a developing country like Uganda?

2.2 The Case of Uganda and its Relationship with the World Bank

2.2.1 Historical Perspectives on Uganda's Governance Discourse

Situated in East Africa, along the equator, Uganda, a British protectorate until its independence on 9th October 1962, was regarded by Sir Winston Churchill as an economic success story and the pearl of Africa (Brett, 1996, Churchill, 2015). At the time of its independence, Uganda had achieved rapid growth in its agricultural sector and enjoyed a vibrant industrial sector. These achievements, adds Brett (1996), were mainly supported by competent economic management and a favourable climate and fertile soils at the time of Uganda's independence from British rule in 1962.

Uganda's GDP expanded at a rate of about 4.2 per cent annually from 1965 to 1971; Brett (1996) adds that this GDP expansion was largely in response to the fast growth of the agricultural sector and its food-processing industries. However, Doornbos (2005) argues that the economic successes and political stability gained during the period as a British protectorate were wiped out by the chaotic rule of Idi Amin Dada from 1971 to 1979, turning Uganda into one of the poorest countries in the world.

Whilst Doornbos (1990) remarked that although Amin's dictatorial rule ended in April 1979 when Tanzanian soldiers invaded Uganda and removed him from power, the country, nevertheless, continued to be embroiled in successive armed conflicts from 1971 to 1986. These endless conflicts and unstable government leadership exacerbated Uganda's political instability and economic ruin.

Makubuya, (1994) reports that Yusuf Lule, for example, who succeeded Idi Amin Dada as president, resigned soon after taking office and was replaced by Geoffrey Binaisa, a former Attorney General of Uganda. Although Makubuya (1994) explains that multi-party elections, which were held in December 1980, returned Milton Obote to power, Makubuya (1994), nevertheless, highlights the fact that dissatisfaction with these election results because of alleged fraud and other election malpractices plunged the country into chaos again. Consequently, multiple guerrilla (rebel) fighting groups emerged, which regularly attacked government forces.

In 1981, General Museveni, who had served in Milton Obote's government as defence minister, and, as a member of the Military Commission, launched a guerrilla war under the auspices of the National Resistance Army (NRA) against Milton Obote's government. Whilst the war was raging between the guerilla forces of General Museveni and of Milton Obote's government, General Tito Lutwa Okello, who was the Army Commander under the Obote government from 1980 to 1985, led a coup d'état that overthrew the Obote regime in July 1985. General Tito Lutwa Okello ruled Uganda as president for six months and was overthrown by the NRA under the leadership of General Yoweri Museveni in January 1986.

The underlying technique that underpinned the successfulness of the organisation of the NRA as a rebel group was mainly based on enlisting the support of the masses, in order to defeat General Tito Lutwa Okello's regime. The philosophical basis for enlisting the support of the masses as a panacea for wining wars was grounded in the Marxist orientation of Mao Zeton (also known as Mao Tse-tung). In this case, the reliance by the NRA on the importance of generating popular support among the masses in order to win this guerrilla war was grounded in the Marxist philosophy of Mao (1975) who argues that "the richest source of power to wage *and win* war lies in the masses of the people" (p.232).

While substantiating Mao Tse-tung's notion on the essential role of the masses in winning guerrilla warfare, Scott (1970), explains that guerilla warfare can also be regarded as a political contest, which is allied to military operations. Scott (1970) adds that the political contest, which is allied to military operations, is, however, an intricate search for a combination of resources that will culminate in victory. Nevertheless, it is contestable as to whether or not the involvement of the masses is essential in winning guerrilla warfare. Whilst Davidson (1981), for example, asserts that the unity, which brings success, is the unity, which achieves participation, Kriger (1992), argues that guerrilla war cannot necessarily be won with the involvement of the masses. Kriger (1992) suggests that guerrilla warfare can, mainly, be won by coercion to achieve the cooperation of the masses.

However, it can be added that the contestation between whether or not the NRA could have won its guerrilla warfare without coercing the masses towards obtaining induced cooperation could have also depended on the extent to which the NRA was committed to its own pursuit for military victory by generating the support of the masses as a means to defeat the government of Milton Obote. Kasfir (2002) notes that in their strategy to win the support of the masses, the NRA leaders identified the failure of the ruling elite to focus on the interests of ordinary Ugandans, coupled with their selfish personal development, as the underlying cause of disorder and dictatorship in Uganda.

Therefore, the fundamental argument advanced by the NRA leaders as the basis for launching a guerrilla war against the Obote government was that, instead of the Obote government focusing on development to achieve the substantive interests of the people, NRA's Ten Point programme (1985) for rebuilding a post-Obote Uganda postulates that post-independence Ugandan politicians and military officers, such as Milton Obote, maintained power by exacerbating sectarian divisions on the basis of religion, region and ethnicity, and based on this, the NRA Ten Point Programme further argues that elections that had been conducted in Uganda, based on these divisions, diverted popular attention away from pressing problems of poverty and lack of development.

In January 1986, General Yoweri Museveni, leader of the NRA overthrew General Tito Okello's regime. Since the overthrow of this regime in 1986 until 1996, political parties and related activities were banned in Uganda, and elections only took place with candidates not affiliated to any political party (Barkan, 2011). In addition, while a new constitution, which re-introduced multi-party politics, was enacted in 1995, the first multi-party elections took place in 2006, and other presidential and parliamentary elections took place in 2011 and 2016, all resulting in an overwhelming win by General Yoweri Museveni who would remain in power until 2021 and become Uganda's longest-serving military leader. Tangri and Mwenda (2010) outline various reasons for General Yoweri Museveni's determination to hold on to power. The belief by General Museveni that he is indispensable to Uganda's stability and prosperity, Tangri and Mwenda (2010) point out, is one of the underlying causes for General Yoweri Museveni's determined grip on power given the history of a country, which has been devastated by decades of bad leadership.

Notwithstanding the persistent grip on power by General Yoweri Museveni, the government of Uganda has opened up space for multi-party elections since 2006. This multi-party dispensation has been followed by various episodes of elections. However, Kifle (2015) shows that whilst the opening up of the democratic space has been consolidated in Uganda, residues of authoritarianism still remain. In addition, Barkan (2011) shows that Uganda's system of governance is "in essence an authoritarian patronage-based regime" (p. 2). Since governance is defined by authors, such as Bevir (2008), Grote and Gbikpi, (2002) and Fukuyama (2013), as the manner in which the State generally acquires and manages public goods and services, Tangri and Mwenda (2005) explain that, like in most African countries, the use of the State and its resources constitute vital forms of consolidating power for Africa's rulers.

The architecture of governance in Uganda, under an authoritarian patronage-based regime, is characterised by what Diamond (2008) describes as an insidiously overpowering presidency which is sculptured around patronage, grounded in ethnic bonding and punctuated by a hierarchical network of patron-client form of state organisation. Diamond (2008) further argues that the organisation of states around autocratic personal rulers, highly centralised and overpowering presidencies and deeply hierarchical, informal networks of patron client relations that draw their symbolic and emotional glue from ethic bonds is assuming greater dominance in African politics.

In summary, I argue that the organisation of African states around autocratic personal rulers, highly centralised and overpowering presidencies seem to generate the misguided impression that African leaders require to exercise extensive capabilities of control and dominance of both the economy and politics in order to organize the necessary resources which are needed for national development.

However, the assumption of such extensive powers, consequently, undermines transparency, accountability, weakens citizen participation as well as the right to free expression, stifles the rule of law, creates a breeding environment for both corruption and poverty as well as increased incidences in violence all of which have depilating effects on the ability of government to effectively dispense inclusive development to its citizens.

2.2.2 The Basis and Effects of the Involvement of the World Bank in Uganda

Rueger (2003) states that "… conceived during World War II at Bretton Woods in New Hampshire in the United States of America in 1944 to provide development support for post-war reconstruction, the original mission of the World Bank was based on pure economic issues" (p.201).

Whilst the original mission of the World Bank was purely economic, Rueger (2003) adds that the Bank has attempted, through its expanded mission, to assist poor countries to develop the necessary infrastructure they need to prosper in the World economy. In addition to its original mission of providing development support based purely on economic imperatives, the Bank's additional responsibility is to reduce poverty in developing countries (see Mosley, Harrington and Toye, 1995; Santiso, 2001; Stern and Ferreira 1997, Williamson, 2002; Miller-Adams 2002). The literature also shows that as part of its social responsibility for the well-being of humanity in developing countries, the World Bank has loose affiliation to the United Nations (UN) as a specialised entity in terms of the United Nations Resolution 41/128, as well as the right of people to be free from poverty as stated particularly in Article 11 of the International Covenant on Economic, Social and Cultural Rights (ICESCR, 1966).

However, while the World Bank has demonstrated its social responsibility with regard to the well-being of humanity, particularly in developing countries, Friedman (1970) clarifies that, like an international financial corporation, the "members of the board are not democratically elected" (Friedman 1970, p. 123). The historical accounts also show that with its first loan of $250 million to France in 1947 for post-war reconstruction, the World Bank started to help to rebuild Europe after the war (Salda, 1997).

In re-iterating the commitment of the World Bank to reducing poverty in developing countries, the (then) President of the World Bank from July 1986 - August 1991, Barber Conable, re-affirmed, in 1988, the overall goal of the World Bank namely:

> "The central goal of the Bank is the reduction of poverty. Poverty on today's scale prevents a billion people from having even minimally acceptable standards of living. To allow every fifth human being on our planet to suffer such existence is a moral outrage. It is more: it is bad economics, a terrible waste of precious development resources. Poverty destroys lives, human dignity and economic potential." (World Bank, 1994, p. 99).

With particular reference to Africa, the rationale for the relationship of the World Bank with African countries is well documented. Heyneman (2003) explains that since political authoritarianism had become pervasive after independence in most African countries, State authority became, undoubtedly, omnipresent in every sector of the economy in African countries. In an attempt to reverse this crisis, the World Bank, in the first instance, worked closely with African countries from the 1960s through the provision of development loans to reform African countries as well as their economies.

The reforms, argues Demery (1994), were needed largely because of both defective political and economic systems that African leaders had established, coupled with misguided policies that the leaders pursued in the post-colonial era. The pervasive influence of the State needed to be rolled back so that African economies could move from state-dominated systems to free market-based systems. The transformation of interventionist countries, and their extensive influence of the governments of African countries in development processes since the 1960s, were undertaken through various episodes characterised by periodic shifts in policy interventions. Williamson (1990) explains that in most African countries, the focus of the World Bank in helping countries, such as Ghana, Nigeria and Zimbabwe in the 1960s and 1970s, was on project-specific lending for infrastructural development of roads, dams for generating electrical power, telecommunications and other public goods.

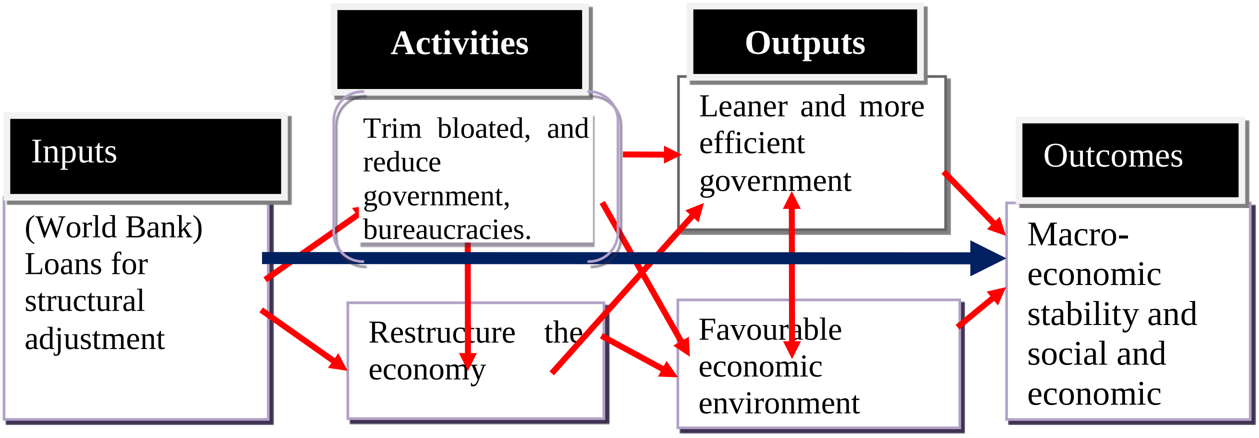
In addition, Whitaker (1988) pointed out that “the World Bank and the International Development Association supplied at least 25 percent of the loans to African countries from the early 1960s” (p.66). Whitaker (1988) further argued that throughout the mid-1980s, the World Bank, the USA and the governments of most African countries felt that development would occur by creating government industries and services that would expand and diversify the economy. However, from 1979 to the early 1980s, the World Bank shifted its focus in African countries from project-specific lending for infrastructure development to structural adjustment lending.

Agyemang (2001, p.1) defined ‘structural adjustment’ as a “process where economic policies and relevant institutions are reformed with a view to enhancing economic growth, improving resource allocation, increasing economic efficiency and the economy’s resilience to changes in its domestic or global markets” .

This shift in the lending policy to African countries, Demery (1994) argues, was a consequence of the emergence of serious economic difficulties in the aftermath of the oil crisis and the downturn in growth rates of the member countries of the OECD. This represented a major external shock for African countries. The loans under this policy regime were extended to the governments of African countries to fund structural adjustment programmes, which included economic re-structuring and poverty eradication (Toye, 1993).

The World Bank, under this policy regime, directed its programmes towards restructuring most of the state-controlled economies towards the private sector and trimmed bloated state bureaucracies into leaner and more efficient governments that could create a more favourable macro-economic environment and lay the foundation necessary for spurring sustainable economic and social development (Toye 1991, 1993). Figure 1 on the next page reflects these assumptions.

Figure 1: Postulated Chain of Results for Structural Adjustment



Source: developed by Author based on the logic in Mayne (1999)

The postulated chain of results in Figure1 demonstrates the assumption that World Bank loans were extended to African governments to fund the restructuring of mostly state-controlled economies towards the private sector and to trim bloated state bureaucracies. This was undertaken in the belief that it would lead to leaner and more efficient governments that create a more favourable environment for macro-economic stability and socio-economic development.

The assumption created in figure1 implies (*through thick line that runs from first box on the left to the last box on the right*) that World Bank interventions lead to macroeconomic stability and socio-economic development. However, whilst the role of government intervention in development in the 1970s and early 1990s was perceived in the context of the World Bank's policy regime, which sought to restructure state controls and trim bloated state bureaucracies as irredeemably unnecessary, in the late

1990s the World Bank started to recognise the role that government plays in economic development. The Bank acknowledged the importance of government in providing goods and services to the public and in designing the rules and regulations that allow markets to flourish (World Bank 1997).

The World Bank affirmed that government also provides important institutional infrastructure, such as laws and regulations that protect property rights and maintain public order, without which sustainable socio-economic development and long-term investment are impossible (World Bank, 1997). Public sector economists, such as North (1981) and De Long and Summers (1993), argue that good institutions, particularly in the public sector, play a vital role in economic growth.

North (1990) additionally proclaims that rules define both the way games are played as well as how they support human interaction. Therefore, an effective institutional infrastructure enhances the performance of economies in the long run, and in addition supports individuals and organisations in taking advantage of opportunities. Effective laws and regulations support the protection of rights as well as enforcing agreements between individuals and organisations. Further, the importance of an effective institutional infrastructure led to the Bank's acknowledgement of the policy shift, which places greater importance on the need to help governments in African countries to improve the planning and implementation capacity through enhancing the capability of public institutions (World Bank, 1997).

The role of good governance in economic development has been emphasised by various authors (Knack and Keefer, 1995; Mauro, 1995; Kaufmann, 2004; Sano, 2002 and Easterly and Levine 1997). Whilst Kaufmann (2004), for example, found a "large direct causal effect from better governance to improved development outcomes" (p.15), nevertheless, Sano (2002) places "increasing weight on good governance in the development discourse" (p.736). In addition, the emphasis on the importance of good governance in development discourse was clearly precipitated by significant evidence of failure by the World Bank with regard to structural adjustment programmes in

African countries in its 1989 report on development. The World Bank's Development in Africa report mentioned failure of governance as a key factor in the failure of economic reforms in African countries in 1989 (World Bank 1989). Therefore, the role played by good governments in fostering economic development in African countries is, indeed, a well-established proposition.

The World Bank report of 2010 showed that since 1960 it had provided more than $400 billion to support poverty-reduction efforts in African countries (World Bank 2010). This report added that, in 1965, almost 20 per cent of the Overseas Development Assistance (ODA) went to Africa and that in the 1980s, Africans, who composed about 12 per cent of the developing world's population, were receiving about 22 per cent of the total ODA, and that the share per capita in Africa amounted to about $20 higher than the $7 allocated per capita in Latin America.

Whilst the World Bank remains committed to the eradication of poverty in African countries, Williamson (2000) argues that within the context of the original mission, which is grounded in the Washington consensus is primarily focused on economic imperatives rather than social imperatives. However, I argue that while the World Bank may seem to re-engineer its engagement in Africa through various periodic shifts in policy; there still exist substantial challenges relating to the ineffectiveness of World Bank poverty reduction interventions in African countries.

Notwithstanding the rationale which underpinned the context of its original mission, the existence of persistent evidence of the ineffectiveness of World Bank poverty interventions in Africa potentially incubates more controversies, and, provides a convenient shield for incompetent African despots to conceal their own failures. African leaders, for example, can find room to argue that there is often no African input in the design and evaluation of development programmes and can also find room to claim, for example, that acceding to World Bank prescriptions for reform amounts to succumbing to foreign dictates. In these circumstances, the World Bank cannot provide an effective framework for combating poverty in African countries.

Nonetheless, in the wake of global changes in the design of global aid in 2001, and following a period of both macro-economic stability and sustainable reform of public institutions in developing countries, the World Bank launched the PRSCs for use in developing countries. The introduction of the PRSCs marked a major departure from the World Bank's adjustment lending model. Compared to the previous adjustment lending model, which imposed inflexible conditions, PRSCs are project-specific and the priority sectors include education and health.

The launch of the PRSCs also followed recognition by the World Bank of the importance of country ownership, the commitment of governments to reform and the recognition of the multi-dimensionality of poverty-reduction efforts. The Bank also used the PRSCs to signal strong support to the commitment of African governments to reform in addition to providing more predictable annual support and strengthening the budget process, all in a results-oriented framework (World Bank 2010). In Uganda, the ruling NRM government espoused governance arrangements characterised by extensive State interventionism.

So the rationale for the involvement of the World Bank through its structural adjustment programmes was mainly to fund restructuring of state-controls, resulting in trimmed and less-bloated State bureaucracies caused by extensive State interventionism. In 1994 Uganda was one of nine African countries (the others being Burkina Faso, Gambia, Ghana, Guinea, Lesotho, Nigeria, Tanzania and Zimbabwe) declared as one of Africa's economic success stories in completing adjustment programmes with increasing returns to sustainable growth and poverty-reduction levels (World Bank,(1994). Uganda continues to receive close attention from the World Bank and attracts generous development loans from the latter towards its poverty-reduction efforts.

However, Mwenda and Tangri (2005) argue that whilst the World Bank hailed Uganda as a major case of economic success in Africa, and provided the government with large amounts of financial assistance to support democratic reforms, such "high levels of foreign aid have provided the government of Uganda with resources to sustain the patronage system of governance" (p. 449). The continuation of the patronage system of governance in Uganda has, in this case, inevitably persistent debilitating effects on efforts to eliminate both corruption and grinding poverty.

Notwithstanding the improvements as well as failures in governance spheres in Uganda over the past 27 years, Amundsen (2006) maintains that Uganda still suffers from various development challenges, such as extreme poverty and high income inequality. In addition, whilst Transparency International, on the one hand, claims that Uganda is highly dependent upon international aid for the delivery of basic services to its citizens, the report also shows that Uganda has a low accountability and transparency rating and widespread corruption, which potentially threaten access to the quality of public services (Milner, Nelson and Findley 2013).

In seeking to address these challenges, the World Bank intervened in Uganda's development processes in order to improve the effectiveness with which the government delivers on its development mandate. Nevertheless, the choice of the World Bank to intervene in the reduction of poverty in Uganda, devastated by several episodes of civil war since its independence in 1962, was mainly precipitated by the role that the Bank had played since its inception in the early 1940s. The historical chronology of the World Bank provides evidence of its involvement in providing development aid to economies in crisis since that time (Mason and Asher, 1973). In addition, a review of the literature of the World Bank showed that while rebuilding countries devastated by calamities, the World Bank today also continues to focus on poverty-reduction in developing economies across the world (Kapur and Webb, 1997).

The UNDP, with specific reference to Uganda, reported in 1980 that the involvement of the World Bank in Uganda's development processes started after many years of conflict lasting from 1971 to 1986, and focused mainly on trimming bloated State bureaucracy so as to improve the socio-economic conditions of the people of Uganda.

In November 1979, the report adds that international agencies, which included the World Bank, held the first international development donors' meeting in Paris to help successive Ugandan governments rebuild their economy after the overthrow of Idi Amin Dada (UNDP 1980). Trebesch, Papaioannou and Uaibir (2012) demonstrate that in June 1981, following the ousting of Idi Amin in April 1979 by Tanzanian-backed soldiers, which brought Milton Obote back into power, the World Bank supported Uganda with a standby facility of US$197 million dollars. This facility marked the starting point at which the Ugandan government came under the support of the World Bank. In 1981, the government of Milton Obote published the economic recovery programme covering the period 1982-1984. Bigsten and Kayizi-Mugerwa (1999) add that this programme focused mainly on addressing pressing social needs as well as boosting the export sector. In addition, a further support programme was announced by the World Bank covering the period 1983-1985, focusing mainly on the industrialisation programme, which covered more than 100 projects (Maehle, Teferra and Khachatryan, 2013).

Furthermore, since the first Paris international donors' meeting in November 1979, the World Bank helped the government of Uganda to adopt economic policies that were directed at stabilisation, rehabilitation and economic recovery following these episodes of conflict. After years of such conflict, which ended in 1986, Uganda adopted the aforementioned economic policies and became one of the few African countries to embrace the stringent structural adjustment programmes, deemed to be critical by the World Bank for restoring fiscal discipline and monetary stability (World Bank, 1994). The reforms undertaken by the Ugandan government endeared the country to the Bank, enabling it to maintain a close-sustained relationship.

While maintaining close attention from the World Bank, as well as attracting generous development loans from the latter towards its poverty-reduction efforts, Uganda also turned out to be a significant advocate for the Bank's programmes in Africa. This country also became the first developing country to complete the debt reduction package and was the first African country to have its PRSP approved, enabling the country to qualify for the enhanced debt relief. Since 1987, the World Bank has provided the Ugandan government with more than US$790 million in loans. In addition, an estimated US$1 billion grants in project support have also been provided to the Ugandan government to support agriculture, infrastructure, export-led industrial growth and social sectors (World Bank 2009). Due to Uganda's commitment to structural adjustment reforms, the country secured a 3year US$63.25 million structural adjustment facility in June 1987. With this funding, Ayittey (2005) remarks that Uganda began undertaking wide-ranging regulatory reforms to streamline the functioning of the economy, aiming at removing and trimming extensive bureaucratic bottlenecks to economic stability and democratisation processes.

In response to these policy and institutional reforms, a major turnaround occurred from the 1990s where Uganda not only attained economic recovery but also achieved sustainable macro-economic stability. Ayittey, (2005) shows that "the rate of inflation fell from 66% in 1986 to 15% in 1993 and below 5% per year for most of the second half of the 1990s annually and also suggests that this robust economic turnaround resulted in the reduction of the rate of interest where it was 240% in 1986 down to 15% in 1993, coupled with an increase in median income per capita from US$200 in 1990 to US$330 in 2000, quite a substantial 65 percent increase" (p.154).

Furthermore, Ayittey, (2005) demonstrates that during the 1990s, Uganda also witnessed "a significant reduction in the incidence of poverty from 56 percent of the total population in 1992 to 35 percent in 2000. The household survey on the incidence of poverty, conducted by the Uganda Bureau of Statistics between 1994 and 1997, shows that Uganda posted a real GDP growth rate of 8 percent, the highest during that period in Africa" (p.154). In addition, the Uganda Bureau of Statistics (UBS)'s integrated Household Survey of 1992/93 demonstrates that "the economic situation continued to improve significantly under the Museveni regime permitting some improvement in the rate of poverty alleviation" (UBS 2004, p.16).

In response, “foreign aid poured in(to) Uganda, amounting in 2000 to some 53 percent of the total government budget, or 13 percent of GDP” (Ayittey 2005, p.154). In addition, Ayittey (2005) showed that “Uganda’s macroeconomic performance also showed an average real growth rate close to 7 percent per year” and adds that “bad weather conditions and transport difficulties caused lower levels of output for coffee and food” (p.16). Whilst the economy was performing well with the population of 20 million in 1995, the influx of refugees from neighboring countries in armed conflict inevitably put Uganda’s health and education services sector under very serious strain, consequently affecting the country’s quality of life in general.

Nevertheless, in spite of the challenges caused by the influx of refugees, Uganda became one of the few African countries to adopt and implement the World Bank’s rigid structural adjustment programmes, which were considered by the Bank as critical pre-requisites for rebuilding the country’s economic discipline and stability. In addition, Uganda became the first developing country to complete the debt reduction package and was the first country on the African continent to have its PES approved by the Bank in 1987.

This approval enabled the country to qualify for the enhanced debt relief, causing the country to be a significant promoter of structural adjustment programmes on the African continent on behalf of the World Bank. Uganda’s robust economic performance in the 1990s also brought a glimmer of hope for sustainable economic stability and growth in the country, resulting in improved continuous relations with the World Bank. As a result, the Bank provided more than “U.S. $790 million in adjustment support, in addition to an estimated U.S. $1 billion in project support in the agriculture, infrastructure, and social sectors” (Ayittey 2005, p176). However, whilst Uganda demonstrated robust economic performance in the 1990s, its external debt stretched to 65 per cent of GDP in 1997, considering that more than 55% of its national budget was mainly funded by the World Bank.

In 2009, the World Bank declared that with this level of foreign debt, Uganda could not sustain high levels of repayment whilst simultaneously achieving high growth rates (World Bank, 2009). Ayittey (2005) demonstrates that in January 1998 Uganda hosted a landmark closed-door meeting between the president of the World Bank, James Wolfensohn, and leaders of twelve African countries to discuss the Bank's policies in Africa. The landmark closed-door meeting between James Wolfensohn and the twelve leaders contributed towards the introduction of the Heavily Indebted Poor Countries (HIPC) initiative which was a "comprehensive strategy intended to examine the entire poverty-reduction expenditure programmes of poor and highly-indebted countries most of them from Africa" (p.153).

In seeking to resolve the external debt crisis, and to benefit from the HIPC initiative, the Ugandan government intensified the transformation of its economy to reduce the incidence of poverty. In order to achieve this and thus benefit from the HIPC initiative, it prepared a Poverty Eradication Action Plan (PEAP) in 1997. The 1997 PEAP provided a comprehensive plan of action the aim of which was to provide direction reduce absolute poverty to less than 10% of the population by 2017 (Ellis and Bahiigwa, 2003).

The 1997 PEAP identified primary health care, rural feeder roads, primary education and provision of safe water, as well as modernisation of agriculture as priority programme areas. The Poverty Action Fund was created by the central government as a vehicle for funding the implementation of these programmes under the 1997 PEAP. Through the Poverty Action Fund, central government provided uniform allocations to local governments, which were charged with the responsibility of implementing poverty-eradication programmes.

However, due to emerging differences in the implementation of the priority areas at local government levels, coupled with the fact that the poor had not been consulted during the formulation of the PEAP 1997, there was a need to refocus some of the issues in the PEAP 1997 in order to enable the Poverty Action Fund effectively on the implementation of the priority areas specified under the 1997 PEAP. Subsequently, the 1997 PEAP was effectively revised, resulting in the PEAP 2000, which also focused on the long-term goal of reducing the incidence of income poverty in Uganda to less than 10% by 2017. Within the context of this goal, the PEAP 2000 committed to "creating an enabling environment for sustainable economic growth and transformation, promoting good governance and security, directly increasing the ability of the poor to raise their incomes and directly improving the quality of life of the poor as its four major priority goals." (Ellis and Bahiigwa 2003, p.100).

The PEAP 2000 was praised as being "homegrown" by its donors, subsequently resulting in situating Uganda as the first beneficiary of the Enhanced HIPC Debt Relief Initiative, and enabling Uganda to access enhanced debt relief under the HIPC Initiative. Ayittey (2005) states that, in April 1998, the World Bank agreed to a US$650 million debt relief package for Uganda, effectively reducing Uganda's external debt by twenty per cent. The impact of the relief was clearly seen on the debt service ratio, in which Uganda exhibited a pre-HIPC figure of US$110 million per year (a 20 per cent debt-service ratio) and a post-HIPC figure of US$35 million per year (a 6-7 per cent debt-service ratio).

In the wake of global changes in the design of global aid in 2001, and, following a period of both macro-economic stability and sustainable reform of public institutions in developing countries, the World Bank launched the PRSCs for use in the latter. The introduction of the PRSCs marked a major departure from the Bank's adjustment lending model. Compared to the previous adjustment lending model, which imposed inflexible conditions, PRSCs aimed to provide more predictable annual support and strengthen the budget process, all in a results-oriented framework (World Bank 2010).

In the context of the PRSCs, the principles of the Paris Declaration on Aid Effectiveness of 2005 are based on the results-based framework (Bissio, 2008). This Declaration commits all countries and organisations in developing countries to continue to increase efforts in managing aid for results with a set of evaluable actions and indicators (Declaration 2005). The launch of the PRSCs also followed recognition by the World Bank of the importance of country ownership, the commitment of governments to reform and the recognition of the multi-dimensionality of poverty-reduction efforts. The Bank also used the PRSCs to signal strong support to the commitment of the Ugandan Government to reform (World Bank 2010).

The PRSCs are project-specific and the priority sectors, among others, include education and health. Uganda was first developing country to complete the debt reduction package and also the first African country to have its Poverty Reduction Strategy approved in 2000, enabling the country to qualify for enhanced debt relief during this same year due to of its strict adherence to the fiscal discipline requirements and its sound economic reform record.

The evaluation of the effectiveness of poverty-reduction interventions in Uganda has been conducted by various organisations since the completion of the debt reduction package, as well as the approval by the World Bank of Uganda's Poverty Reduction Strategy in 2000. The 2005 United Nations Human Development report, for example, showed that more than 20% of Ugandans live below US$1.25 dollar a day and that inequality had worsened where the income difference (measured by the Gini Coefficient) had worsened from 0.33 to 0.43 (UNDP,2005). In addition, the assessment of poverty and vulnerability by the World Bank in 2006 demonstrated an increase in the poverty rate (World Bank, 2006; Sala-i-Martin, 2006). Furthermore, the evaluation of its AAP from 2001 to 2007 adjudged specifically that although Uganda had posted robust economic growth during the same period, these economic growth rates had not, paradoxically, translated into overall reduction in the incidence of poverty (World Bank 2009).

However, sector-specific evaluations of the World Bank's support to Uganda through the PRSCs, specifically to "improve water and sanitation as part of its human development agenda, showed satisfactory performance" (World Bank 2009, p. 17). What could be learnt specifically from selected World evaluation cases, which assessed the effectiveness of the World Bank's poverty-reduction interventions funded through the PRSCs in Uganda?

2.2.2.1 Description of the World Bank Evaluation Cases in Uganda

With scant evidence from Uganda on the utilisation of metaevaluation to assess the quality of evaluation of poverty-reduction interventions, it was shown in the last paragraph of section 2.1.3 of this chapter that metaevaluation of assessing the quality of evaluations of development interventions has been applied mainly in different contexts in developed countries. This study on metaevaluation of evaluations of the World Bank poverty-reduction interventions in Uganda used four completed evaluations with which to assess the quality of the evaluations conducted under the auspices of the World Bank on poverty-reduction interventions in Uganda.

The World Bank evaluates its programmes and activities to ensure the integrity of the Bank's self-evaluation process, to verify that the work of the Bank is producing the expected results and to help develop improved directions, policies and procedures through the dissemination of lessons drawn from experience. In order to prepare a project performance evaluation report, the staff of the World Bank's Independent Evaluation Group (IEG) examine project files and other documents, interview operational staff, visit the borrowing country to discuss the operation with the government and other in-country stakeholders, and interview World Bank Staff and other donor agency staff, both at headquarters and in local offices as appropriate (World Bank, 2009).

The project performance evaluation report of the World Bank is subjected to internal peer review by World Bank's IEG and management approval (World Bank, 2009). Once cleared internally, the project performance evaluation report is commented upon by the relevant department within the World Bank. The IEG incorporates the comments, based on their relevance. The completed project performance evaluation report is then sent to the borrower for review; the borrower's comments are attached to the document sent to the Bank's Board of Executive Directors, after which the project performance evaluation report is released to the public. The four selected completed evaluation cases for this thesis were conducted in order to assess the effectiveness of poverty-reduction interventions under the auspices of the World Bank in Uganda. The selected evaluation cases were on poverty reduction projects funded by the World Bank through PRSCs to the government of Uganda. The selected evaluation cases for this thesis were released to the public in 2009 (World Bank 2009).

The PRCS in Uganda were developed in the context of the World Bank's Country Assistance Strategy (CAS) for the financial year 2001-2005 (World Bank, 2009). The CAS supported Uganda's economic transformation and poverty-reduction strategy as articulated in the government's 2000 PEAP. The PRSCs were used as financial instruments to fund Uganda's poverty-reduction efforts based on the country's PEAP of 2000 (World Bank, 2009).

The first case is an evaluation of PRSC1 under Credit Number 3510, which was approved on 31st May 2001 and became effective on 29th November 2001 and closed on 31st March 2002 (World Bank 2009*a*, p.3). PRSC1 had four objectives. The first objective was to support economic growth and transformation by improving efficient, equitable use of public resources.

The second objective was to ensure good governance and security by improving service delivery through cross-cutting reforms, such as improvement of management systems and practices in the public sector, as well as increasing transparency, participation and reducing poverty. The third objective was to increase directly the ability of the poor to raise their incomes, and the fourth objective was to increase directly the quality of life of the poor by improving the quality of both education and healthcare, and to improve access and achieve equity in water and sanitation.

The second poverty-reduction intervention, funded through the second Poverty Reduction Support Credit (PRSC2) under Credit Number 3696, which continued to support Uganda's Poverty Eradication Action Plan (PEAP) of the year 2000 was approved on 23rd July 2002, became effective on 12th March 2003 and closed on 30th June 2003.PRSC2 continued to support an enabling environment for rural development. PRSC2, unlike PRSC1, included the third goal of Uganda's PEAP 2000, which was to increase the ability of the poor to raise their incomes. PRSC2 had four main objectives (World Bank, 2009*b*, p. 13).

In order to support the implementation of the framework for economic growth and transformation, the first objective of PRSC2 was to enhance the efficient and equitable use of public resources. The second objective was to improve service delivery through cross-cutting reforms; it supported the second goal of Uganda's PEAP, ensuring good governance and security. In seeking to increase directly the ability of the poor to raise their incomes, the third objective of PRSC2 was to promote an enabling environment for rural development.

The fourth objective sought to support the effective implementation of Uganda's PEAP of 2000, that is, directly increasingly the quality of life of the poor by improving the delivery of basic services. The fourth objective of PRSC2 focused on improving the quality of education and healthcare, and improving access and achieving equity in water and sanitation.

PRSC3, under Grant Number H0630, was approved on 9th September 2003, became effective on 11th February 2004 and closed on 30th September 2004. PRSC3 supported the implementation of the four goals of Uganda's PEAP. PRSC3 continued to support the PEAP through the first and second PRSCs with similar objectives (i.e. on improving the quality of education and healthcare, and improving access and achieving equity in water and sanitation) (World Bank *c*, 2009, p. 21).

PRSC4 under Grant Number (H1220-UG) was approved on 2nd September 2004, became effective on 28th February 2005, and closed on 30th September 2005. Although PRSC4 continued the World Bank's support to Uganda for its service delivery and rural development reform programme, the main thrust of PRSC4 was to support the implementation of all the specific objectives for PRSC1, PRSC2 and PRSC3 further (World Bank, 2009*d*, p. 27).

Therefore, it was deduced that the individualised PRSCs, which Uganda received from the World Bank at various successive intervals from 2001 to 20015 to support its poverty-reduction efforts, resulted in steady progress in the effectiveness of poverty reduction interventions. The next section scrutinises evaluation outcomes of selected evaluation cases of poverty-reduction interventions funded through the World Bank's PRSCs in Uganda in order to determine evidence of inconsistencies in the evaluation outcomes of the effectiveness of selected poverty reduction interventions.

2.2.2.2 Scrutiny of Evaluation Outcomes of Selected Evaluation Cases.

In determining evaluation outcomes of the effectiveness of development interventions, the World Bank uses three evaluation dimensions to assess the extent to which an intervention has achieved its stated objectives. These evaluation dimensions are relevance, efficacy and efficiency. Relevance assesses the importance of objectives and of designs. Relevance of objectives is the extent to which the project's objectives are consistent with a country's development priorities (in this case poverty-reduction priorities in accordance the country's poverty-reduction strategy papers), and, relevance of design is the extent to which the project's design is consistent with the stated objectives.

Efficacy is the extent to which the project's objectives were achieved and efficiency is the extent to which the project achieved a return higher than the opportunity cost of benefits at the least costs compared to alternatives. In terms of the World Bank, evaluation outcomes of poverty-reduction interventions are rated as follows: "Highly

Satisfactory, Satisfactory, Moderately Satisfactory, Moderately Unsatisfactory, Unsatisfactory and Highly Unsatisfactory" (World Bank, 2009, p. 3). Below is a summarised scrutiny of evaluation outcomes of selected World Bank evaluation cases.

First Poverty Reduction Support Credit (PRSC1*)*: Objectives and Evaluation Outcomes.

The first poverty-reduction intervention funded through the World Bank's PRSC1 under credit Number (3510) had the following objectives from 2001-2002: "improving efficient use of Public Resources, improve service delivery through crosscutting reforms such as improving management systems and practices in the public sector, increase transparency, participation and reduce corruption; improving delivery of basic services such as the improvement of the quality of education, healthcare and access and equity in Water and Sanitation" (World Bank 2009*a*, p. 3).

The evaluation outcomes for the first Poverty-Reduction Support Credit (PRSC1) demonstrate that the intervention "helped in increasing access to education, water supply and sanitation services. Evidence on budget allocations and executions under PRSC1 indicated that the execution fell below the allocations for sectors, such as education, health, agriculture and administration that could have helped to raise productivity. In addition, anticipated impacts of crossing-cutting reforms did not materialise as, in some cases, the reforms were not carried out and the evaluation outcomes rating was moderately satisfactory" (World Bank 2009*a*, p. 9).

Second Poverty Reduction Support Credit (PRSC2*): *Objectives and Evaluation Outcomes.

With the exception of the promotion of an enabling environment for rural development as an additional objective under PRSC2, the latter, under credit Number (3696) continued to implement similar objectives as stipulated under PRSC1 from 2002-2003 i.e. [see paragraph above] (World Bank 2009*b*, p.13). The evaluation outcomes for PRSC2 from 2002 to 2003 demonstrated that the interventions "helped to achieve the expenditure targets partially in increasing access to health, and water and sanitation services, and in a number of classrooms, reducing the number of students per classroom.

In terms of rural development, the evaluation outcomes demonstrate an increase in the investment budget for agriculture, a necessary and important step for generating and disseminating technology, but revealed that such increase was insufficient to change living conditions in the short-term. Evaluation outcomes under PRSC2 also show that, despite the World Bank’s commitment to reforms aimed at increasing transparency, participation and reduction of corruption since the implementation of PRSC1, little changed in terms of transparency and reduced corruption, and the evaluation outcomes rating was only moderately satisfactory” (World Bank 2009*b*, p.18)

Third Poverty Reduction Support Credit (PRSC3): Objectives and Evaluation Outcomes.

Under grant Number (H0630) PRSC3 continued to implement all the objectives pursued under PRSC2 “improving efficient use of Public Resources, improving service delivery through cross-cutting reforms such as improving management systems and practices in the public sector, increase transparency, participation and reduce corruption; improving delivery of basic services such as the improvement of the quality of education, healthcare and access and equity in Water and Sanitation, promotion of an enabling environment for rural development” (World Bank 2009*c*, p.21). The outcomes of the evaluation of PRSC3 demonstrate that the interventions “did not lead to more efficient and equitable use of public services nor to the cross-cutting reforms that could improve service delivery.

However, the evaluation outcomes show that the interventions helped to increase access to education, health, and water supply and sanitation. Nevertheless, the evaluation outcomes reveal that there was a failure to link outputs and procedures to anticipated outcomes, and that the evaluation outcomes rating was only moderately satisfactory” (World Bank 2009*c*, p.26)

Fourth Poverty Reduction Support Credit (PRSC3): Objectives and Evaluation Outcomes.

Under grant number H1220-UG, PRSC4 continued to implement all the objectives stipulated under PRSC1, PRSC2 and PRSC3 "improving efficient use of Public Resources, improve service delivery through cross-cutting reforms such as improving management systems and practices in the public sector, increase transparency, participation and reduce corruption; improving delivery of basic services such as the improvement of the quality of education, healthcare and access and equity in Water and Sanitation, promotion of an enabling environment for rural development" (World Bank 2009*d*, p.27).

The outcomes of the evaluation of PRSC4 reveal that the interventions "did not lead to more efficient and equitable use of public resources nor to cross-cutting reforms that would improve service delivery, but helped to increase access to education, health, and water supply and sanitation services, however, the evaluation outcomes rating was moderately unsatisfactory" (World Bank 2009*d*, p.31).

Table 1 below provides a summary of Evaluation Outcome Ratings based on each selected evaluation case.

Table 1: Ratings of Evaluation Outcomes Based on Each Evaluation Case

Poverty Reduction Interventions	*Evaluation Outcomes Ratings*
PRSC1 (Credit No.5310) from 2001-2002	Moderately Satisfactory
PRSC2 (Credit No.3696) from 2002-2003	Moderately Satisfactory
PRSC3 (Grant No.H0630) from 2003-2004	Moderately Satisfactory
PRSC4 (Grant No.H1220-UG) from 2004-2005	Moderately Unsatisfactory

Source: World Bank (2009*a*).

2.2.2.3 Summarised Results from the Scrutiny of Evaluation Outcomes

The scrutiny of evaluation outcomes in general revealed evidence of remarkable inconsistencies in evaluation outcomes in the effectiveness of all four selected poverty-reduction interventions of the World Bank in Uganda during the period 2001 to 2005. In addition, the summary of evaluation outcomes of the first three selected evaluation cases in Table 1 provided glaring evidence of worsening evaluation performance rating from moderately satisfactory to moderately unsatisfactory from the period 2001 to 2003 and from 2004 to 2005 respectively. In particular, the least rating (moderately satisfactory) in evaluation outcomes in the effectiveness of poverty-reduction interventions was during the periods 2001 to 2002, 2002 to 2003, 2003 to 2004 to the worst rating (moderately unsatisfactory) in evaluation outcomes in the effectiveness of the fourth selected evaluation case during the period 2004-2005.

In practice, development funders are mainly interested in steady consistency in the effectiveness of their funding support. However, it was particularly clear from these evaluation outcomes, in terms of the scrutiny of general outcomes as well as the ratings that there was evidence of low-grade outcomes in the effectiveness of poverty reduction interventions.

The evidence from the assessment of evaluation outcomes, for example, showed that expected impacts of crossing-cutting reforms did not materialise since, in some cases, the reforms were not carried out, for example in the case of PRSC1. In terms of rural development outcomes, the evidence from evaluation outcomes revealed that the increase in the investment budget for agriculture, which was deemed as a necessary, important step for generating and disseminating technology, was insufficient to change living conditions in the short-term, and little changed in terms of transparency and reduced corruption in the case of PRSC2.

In the cases of PRSC3 and PRSC4, evaluation outcomes in the effectiveness of poverty-reduction interventions revealed that the interventions did not lead to more efficient and equitable use of public services nor did they lead to the cross-cutting reforms that could improve service delivery. These inferior evaluation outcomes in the effectiveness of poverty-reduction intervention in the case of PRSC1, PRSC2, PRSC3 and PRSC4 were inconsistent with the satisfactory evaluation results of the World Bank's poverty-reduction interventions in specific sectors, such as health, economic growth, public reform, agriculture, education, and water and sanitation in Uganda across the four selected evaluation cases.

The dynamic assortment of evident inconsistencies in evaluation outcomes in the effectiveness of selected evaluation cases of poverty-reduction interventions in Uganda, as well as the demand by development funders for steady consistency in the effectiveness of their funding support, raised serious questions about the legitimacy and quality of these selected evaluations, as well as the evidence used to support decision-making.

The questions about the legitimacy and quality of these selected evaluation questions implicitly heightened the importance of evaluating the quality of these evaluations of poverty-reduction interventions. The evaluation literature promotes the importance of evaluating evaluations in order to determine the legitimacy and quality of evaluation evidence. Cooksy and Caracelli (2005) claim that evaluation results influence decision-making processes that are aimed at improving the effectiveness of social interventions, and that the capability for continuous improvement of social interventions can be undermined when evaluations are flawed.

In order to improve the effectiveness of social development interventions consistently, concern with evaluating the merit and worth of evaluations of social development interventions was of paramount importance. The evident inconsistencies in the evaluation outcomes, coupled with scant evidence in Uganda, which identified and resolved the causes of these inconsistencies, raised serious questions about the overall legitimacy and justifiability of evaluation evidence used to support judgments on the effectiveness of these poverty-reduction interventions in Uganda. What then could be learnt from the metaevaluation of evaluations of poverty reduction interventions funded through the PRSCs by the World Bank in Uganda’s case study?

2.3 Conclusions

Situated in the clarificative form of evaluative enquiry, metaevaluation (evaluation of evaluations) is a unique and specialised form of evaluation, which is grounded in the theory for good social programme evaluation and idealised within the context of solving social problems faced by society. Metaevaluation helps in judging the strengths and weaknesses of primary evaluations and enhances the confidence of the users of the results of the primary evaluation, so that interested parties are able to judge the reliability and credibility of evaluations and the evidence used to make decisions.

Metaevaluation has been utilised in different contexts in some developed countries to assess the quality of evaluations of social development interventions; however, the evidence in the literature of evaluation regarding the utilisation of metaevaluation to assess the quality of evaluations of development interventions had remained scanty in developing country contexts until I embarked on this study, particularly in the case of Uganda. After many years of conflict, which ended in 1996, Uganda adopted governance policies directed at stabilisation, rehabilitation and recovery, which endeared the World Bank to provide the Ugandan Government with funding through an assortment of PRSCs to support the country's poverty-reduction efforts.

In practice, development funders of poverty-reduction interventions are mainly interested in steady consistency in the effectiveness of their funding support. A scrutiny of evaluation outcomes of the effectiveness of the four selected World Bank poverty-reduction interventions, funded through its (PRSCs in Uganda, however, revealed evidence of remarkable inconsistencies in evaluation outcomes in the effectiveness of selected poverty-reduction interventions of the World Bank in Uganda, which were funded at successive annual intervals from 2001 to 2005. These inconsistencies in evaluation outcomes over the four-year period raised serious questions about the legitimacy and quality of the evaluations of these selected poverty-reduction interventions.

The questions regarding the legitimacy and quality of these evaluations, as well as the demand by development funders for steady consistency in the effectiveness of their funding support reinforced the importance of evaluating the quality of such evaluations in order to identify and resolve the causes of such inconsistencies. The methodology for evaluating the quality of four selected evaluations of poverty reduction interventions funded by the World Bank through the PRSCs in order to identify and resolve evident causes of such remarkable inconsistencies in evaluation outcomes of selected evaluation cases, as well as answering the specified research question, are outlined in the next Chapter Three.

CHAPTER 3: RESEARCH METHODOLOGY

3.0 Introduction

This study used the mixed methods approach composed of a case study approach, document analysis and descriptive frequency statistics. The use of mixed methods allowed for triangulating a multiplicity of selected metaevaluation methods, which provided multiple sources of reliable evidence. The mixed methods approach supported the process of triangulating multiple perspectives and information sources to generate enhanced credibility of evidence.

The use of mixed methods also insulated the study from the generation of evidence based on what could be evidently construed as pieces of evidence and probable biases associated with a single methodological approach. The philosophical basis for the methodological approach for this study was grounded in the constructivist paradigm and situated within the constructivist evaluation. The preceding chapter situated metaevaluation in the clarificative form of evaluative enquiry and determined evident inconsistencies in the nature of evaluation outcomes in the effectiveness of selected World Bank poverty-reduction interventions in Uganda's case study. The object of this chapter is to specify and discuss the methodology for data collection and analysis so that the underlying causes of these evident inconsistencies could be clarified.

The chapter is divided into four major sections. The first section provides the philosophical basis within which the research methodology is grounded. Section Two presents the rationale for a case study approach. The use of document analysis and descriptive statistics is discussed in Section Three. Section Four provides conclusions from the chapter.

3.1 Philosophical Basis for the Methodology

Since metaevaluation (evaluation of evaluations) is a specialised form of evaluation, the methodological approach for this study was also situated within the context of evaluative enquiry, and based on constructivist evaluation which is grounded in the constructivist philosophical paradigm.

Evaluation is one of the three basic disciplines of enquiry, the others being policy analysis and research (Lincoln and Guba, 1985). Evaluation is a form of enquiry, which focuses on the effectiveness of the evaluand. The evaluand is the subject of an evaluation and can be a project, programme, organisation or process. The evaluation of the evaluand results in merit and worth judgments/constructions (Guba and Lincoln, 1989, Guba and Lincoln 2001). Merit constructions/judgments focus on the intrinsic quality of the evaluand and worth judgments/constructions focus on the extrinsic usefulness of the evaluand (Guba and Lincoln 2001).

Constructions on the intrinsic quality of the evaluand focus mainly on determining the extent to which the project, programme, organisation or process is relevant to a specific context, and constructions on the extrinsic usefulness focus on determining the outward beneficial impact of the project, programme, organisation or process to a specific localised context. Guba and Lincoln (2001) define constructivism as "a theory of knowledge that argues that humans generate knowledge and meaning from interactions between their experiences and their ideas" (p.2), whereas, constructivist evaluation is a form of enquiry, which results in constructions and judgments on the merit and worth of a specific evaluand, such as a project, programme, organisation or process.

The constructivism paradigm is based on three fundamental assumptions of knowledge. These assumptions of knowledge include ontological assumptions of relativism, epistemological assumptions of subjectivism and methodological assumptions of hermeneutic-dialecticism (Guba and Lincoln 2001). Ontological assumptions of relativism propagate the claim that any point of view does not imply absolute truth and therefore, such a point of view cannot be claimed to assume monolithic validity. For example, descriptions of factual evidence on the nature and effects of the project, programme, organisation or process are not presumed to represent absolute truth.

Baxter and Jack (2008) explain that the constructivist assumptions of relativism postulate the relativism of truth. Epistemological assumptions of subjectivism advance the claim that knowledge is subjective and dependent upon both one's experiences and one's ideas. The process of collecting and interpreting knowledge about the nature and effectiveness of the project, programme, organisation or process, therefore, culminates in subjective knowledge.

Methodological assumptions of hermeneutic-dialecticism promote the understanding that the process of the systematic assessment of the nature and effectiveness, as well as the interpretation of resulting knowledge from the evaluand, such as the project, programme, organisation or process, recognises the importance of the involvement of various interested parties, as well as the usefulness of the triangulation of tools of investigation and interpretation in constructing knowledge. However, Stige, Malterud and Midtgarden (2009) clarify that although the 'constructivist paradigm recognises the importance of the subjective human creation of meaning; it does not reject the notion of objectivity. Therefore, it can be inferred based on this postulation on the importance of the subjective human creation of meaning, that objective views on reality can be described through the case study approach.

3.2 The Rationale for a Case Study Approach

The four selected cases for this study were discussed in detail in Sections 2.2.3 and 2.2.4 of this thesis. Yin (2009) defines a case study as an "empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context" (p.18). Yin (2009) urges that a case study method emphasises detailed contextual analysis of conditions and their relationships. In addition, Stake (1995) elaborates that a case study captures the particularity and complexity of a single case with the aim of understanding its activity within important circumstances. Yin (2009) also emphasises that a case study approach provides the basis for systematic testing with a larger number of cases. Stufflebeam and Shinkfield (2007) state that a case study approach is highly appropriate in programme evaluation because it looks at the programmes as they naturally occur and addresses accuracy issues by employing and triangulating multiple perspectives and information sources.

Gummesson (1988) explains further that the case study approach enables us to study many different aspects and to examine them in relation to each other. Lamnek (1995) also clarifies that a case study is appropriate for the analysis of a single phenomenon against the background of its context with the aim of presenting the relevant situations as objectively and as free from prejudice as possible. In general, Stake (1995, 2006), Yin (1989, 1993, 2009) and Meyer (2001) state that single-case and multiple-case studies are the two major categories of case study research methodology.

Yin (2009) adds that a "single case study is analogous to a single experiment which must represent a critical, unique, representative case in testing a well formulated problem issue" (p.47). Furthermore, Yin (2009) observes that a single case study is revelatory in that an "investigator has an opportunity to observe and analyze a phenomenon previously inaccessible to social inquiry, as well as studying such a case at two or more different points in time."(p.48).Therefore, it is very clear from the preceding review that the closeness of a single-case study to real life situations provides a wealth of details and useful strands of knowledge.

Nevertheless, there are observations in the literature, which indicate that knowledge gained from a single-case study cannot be generalised. Giddens (1984), for example, explains that research, which is geared towards explaining the problems of society can be generalised if such research cuts across a wide range of action contexts. The evident limitation in the generalisability of the findings of a single-case study is a major impediment to its full acceptability. Barton (1990) suggests, therefore, that a multiple-case approach is one way to resolve the evident discontent with a single-case study approach and explains that a variety of case studies enhances the credibility of evidence which is used to inform interpretations.

3.3 Document Analysis and Descriptive Statistics

Document analysis formed the major component of the dimension for data collection and analysis, and descriptive statistics formed the minor component of the dimension for data analysis. Document analysis is defined by Bowen (2009) as a "systematic procedure for reviewing or evaluating documents" (p.27). Bowen (2009) explains that the systematic analytical procedure involves identifying, choosing, valuing and integrating data obtained from documents. In addition, Corbin and Straus (2008) and Rapley (2007) explain that document analysis requires that data be scrutinised and clarified in order to extract meaning, gain understanding and develop empirical knowledge. Rossman and Wilson (1995), for example, used "reviews of documents as primary sources of qualitative data to evaluate the effectiveness of Regional Educational Agencies." (p.636). In the case of document analysis, the content of each selected evaluation case for this study was assessed using the quality checklist stipulated in the PES, based on a set of quantitative scales using the four selected completed evaluation reports/cases on poverty-reduction interventions in Uganda funded by the World Bank through its PRSCs between 2001 and 2005 as discussed in Sections 2.2.3 and 2.2.4 of this study. The process of data collection and analysis was limited to an examination of the quality of the four selected evaluations as above using the quality checklist and prescribed quantitative scales stipulated in the PES (The Joint Committee (1994); Sanders (1994); See summarised list of checklists, scales and scores in Appendices A to E, p.196-213).

The tasks involved reviewing evaluation goals, merit, designs, assessment of the credibility of the evaluation team members and examination of their methods of data collection and analysis specified in each selected evaluation case. The tasks also involved an assessment of the systematic nature of the evaluation and social responsibility of the process of designing and evaluating selected interventions. The judgments regarding the adequacy and strength of each selected evaluation case in meeting the requirements of a good evaluation stipulated under the PES were made based on the following set of prescribed quantitative scales: "0-2 Poor, 3-4 Fair, 5-6 Good, 7-8 Very Good, 9-10 Excellent". Literature on metaevaluation recommends that "an evaluation be failed according to the PES if it scores 'Poor' on the following Standards: P1 Service Orientation, A5 Valid Information, A10 Justified conclusions or A11 Impartial Reporting" (Stufflebeam, 1999, p.2; see also Appendices C and E: p.202 and p.213).

Descriptive statistics which formed the minor component of the dimension for data analysis were used to summarise the data, which was generated from document analysis using numerical scores based on prescribed quantitative scales stipulated under the PES. Using descriptive statistics, quantitative scores generated from a set of prescribed quantitative scales based on the checklist prescribed by the PES for individual evaluation standards and associated criteria were subjected to frequency analysis using the SPSS. Summarised descriptive frequency data outcomes were used to support in simplifying the meaning in the description and interpretation of emerging frequency patterns.

3.4 Limitations of the research and Validity of the findings

The study used four selected evaluation cases and was limited to the assessment of the quality of each of the four selected evaluation cases funded by the World Bank through its PRSCs between 2001 and 2005 in Uganda using the evaluation quality checklist and quantitative scales prescribed under the PES. The prescribed quantitative scales and scores based on the quality checklist prescribed under the PES was used for assessing the quality of these selected evaluations (See also specifically Appendices A-E, (p.196-213). The contents of the selected four evaluations cases were analysed using the checklist prescribed under the PES to generate quantitative scores.

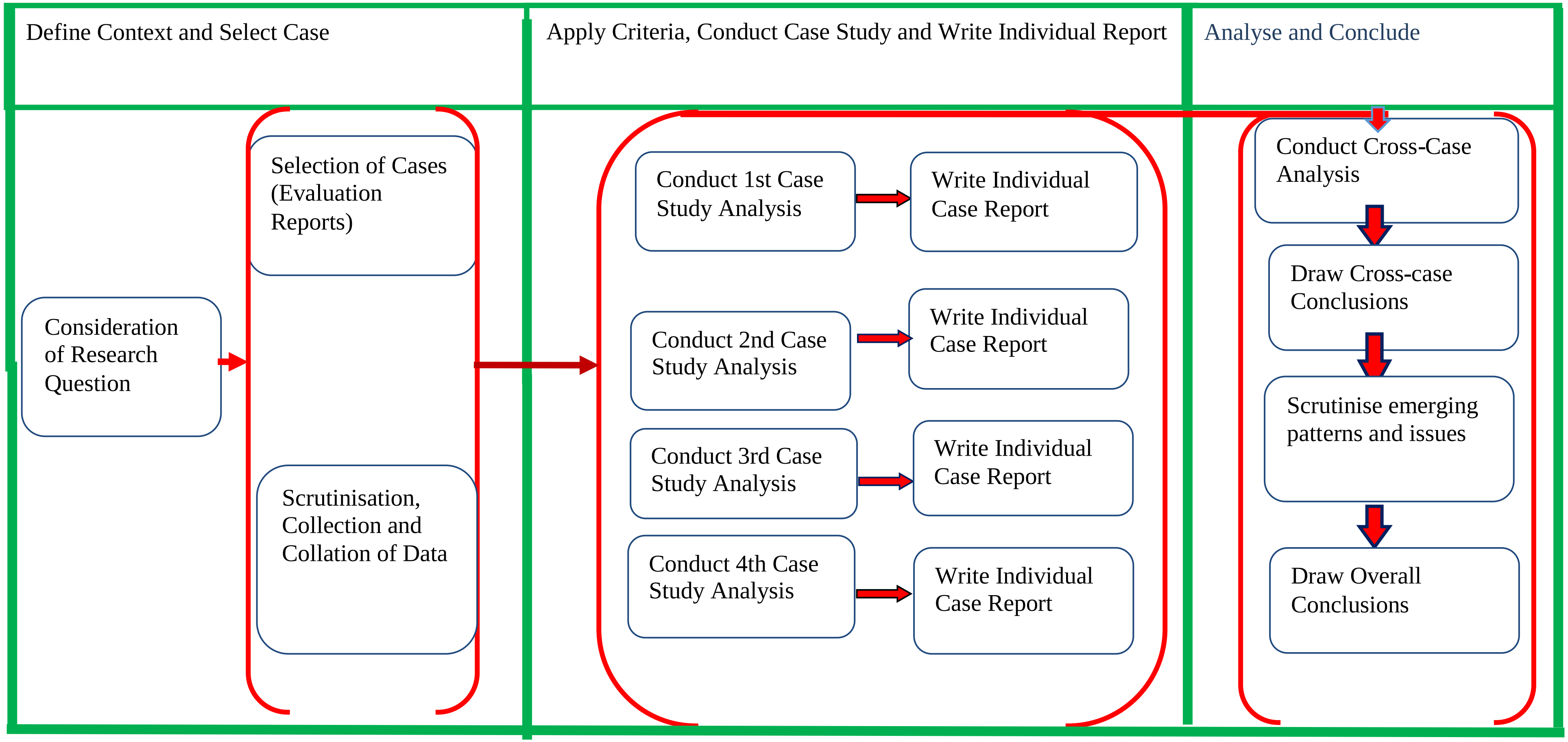
The quantitative scores from a set of prescribed quantitative scales based on the checklist prescribed by the PES were subjected to descriptive frequency statistics scrutiny using SPSS to generate and identity frequency patterns across selected evaluation cases, across the checklists prescribed under the PES and at the level of individual evaluation standards. The emerging descriptive frequency statistics patterns assisted in generating new insights into the outcomes from emerging observations and creating meaning based on emerging patterns.

Within the context of the constructivist paradigm, it was discussed under section 3.1 of this chapter 3 that humans generate and interpret knowledge and meaning from their investigative endeavours based on their interactions, experiences and their ideas within a specified societal context. The researcher argues that within the context of the constructivist paradigm, determinations on the quality of each selected evaluation case was based on an assortment of different forms of human investigative behaviours which underpinned knowledge constructions and judgements on the evaluation outcomes of each selected evaluation case.

Therefore, while judgements on the quality of each evaluation case reflected a variability of constructions based on different forms of investigative behaviours, the triangulation of multiple perspectives and information sources generated from the use of the mixed methods approach for this study enhanced the validity of the findings. The importance of the effectiveness of triangulating multiple perspectives and information sources is reiterated by Eisner (1991) who emphasises that the use of multiple sources of evidence generates “a confluence of credibility” (p.110). In this case, the process of triangulating multiple perspectives and information sources enhanced the credibility of evidence.

Figure 2 on the next page demonstrates the sequence that was followed in the application of the tools of data collection and analysis for this metaevaluation study.

Figure 2: Multi-Case Metaevaluation Application Model



Source: Based on the logic in Yin, R.K (2009, p.57) but contextualised by author.

Regarding multi-case analysis as reflected in Figure 2, Stake (2006) argues that the cases for multi-case studies need to be similar. In this study, whilst allowing for a contrast between cases through a deeper scrutiny of each case, the four selected evaluation cases were similar and supported generalisability more than a single evaluation could. In terms of the multi-case application loop specified in Figure 2, discovery of knew knowledge occurred during the assessment of the quality of each selected evaluation case at the analysis and conclusion stage.

3.5 Guidelines for Conducting Metaevaluation

Evaluation literature proposes a variety of key guidelines for conducting metaevaluation. Stufflebeam (1974), for example, proposes that metaevaluation should "assess the merit; serve the decision-making and accountability imperatives; assess goals, designs, implementation and results; provide judgmental information and appropriate recommendations; serve all persons who are involved in and affected by the evaluation; be conducted by both insiders and outsiders; be a process of delineating the questions to be addressed; obtaining the needed information and using the information in decision-making and accountability; and be technically adequate." (p.67); (See also Appendix E, p.213).

Patton (1997) specifies the questions on which a metaevaluation exercise must focus, such as: Was the evaluation well done? Is it worth using? Did the evaluation meet professional standards and principles? Patton (1997) further recommends that metaevaluation should be conducted with consideration of political contexts.

In addition, Scriven (2007) recommends the use of the evaluation checklist or standards of the PES developed by the Joint Committee for Program Evaluation of 1994. Wholey, Hatry and Newcomer (2004) state that an evaluation must entail a description of the resources, activities, process and results caused by the programme, and identify issues, such as the relevance, effectiveness, efficiency and sustainability of the programme.

In addition, Bollen, Paxton and Moroshima (2005) suggest that metaevaluation should focus on three necessary ingredients, namely: information on inputs, information on results and controls for confounding factors. These guidelines were used for assessing the quality of evaluations and served to determine the degree to which an evaluation complied with acceptable requirements based on the PES (see a detailed checklist for conducting metaevaluation is provided in Appendices A-D, p.196-205.

3.5.1 Standards and Criteria for Metaevaluation

Fitzpatrick, Sanders and Worten (2004) state that there must be evidence of a broad shared consensus on the factors that constitute a good evaluation. Fitzpatrick, Sanders and Worten (2004); Patton (1997) and Russ-Eft and Preskill (2001); Stufflebeam (2001); Stufflebeam and Shinkfield (2007) and Wingate (2010), generally, demonstrate that the PES meets this criterion in terms of broad shared consensus in term of the factors that constitute a good evaluation.

The PES is a set of generally-accepted standards of quality evaluation, which was developed by the Joint Committee for Program Evaluations in 1994 of the American Evaluation Association (AEA) (The Joint Committee, 1994). Beywl and Speer (2004) affirm that the PES is useful as a framework for assessing the quality of a particular evaluation.

Even beyond the USA's boundaries, evaluation associations in other countries have also confirmed the utility of the PES for improving evaluation practice. For example, the German and Swiss Evaluation Societies as well as the African Evaluation Association (DEGEVAL, SEVAL and AfrEA) have respectively adopted a modified, contextualised version of the United States' PES.

The Joint Committee for Program Evaluation of 1994 maintains that the PES promotes consistency in practice and diminishes the potential for evaluations to be judged on personal criteria that do little to advance the evaluation profession or educate evaluation consumers as to the factors that constitute a good evaluation (Sanders, 1994). The joint committee further declares that the PES assists in preventing judgments about the quality of evaluations that centre on methodological or theoretical preferences, personal opinion or arbitrary criteria. Stufflebeam (2001) explains that for operational purposes, metaevaluation should involve the "process of delineating, obtaining and applying descriptive and judgmental information concerning the following four categories of evaluation standards: utility, feasibility, propriety and accuracy; and their systematic nature, competence, integrity, respectfulness and social responsibility to guide the evaluation and publicly report its strengths and weaknesses" (p.186).

Stufflebeam (2001) further explains that the "process tasks of delineating and applying" mean assessment of the participation of stakeholders of the evaluation. The obtaining elements were the technical tasks required to acquire and assess the information needed to judge the target evaluation. Involved are the collection of information through the assessment of evaluation contracts, plans, instruments, data and reports (Stufflebeam 2001, p.186). The Joint Committee for Programme Evaluation of 1994 (The Joint Committee, 1994) provides 30 standards under four major categories. A condensed taxonomy of the key elements of each of the four major categories for checking the merit and worth of evaluations is outlined individually below. A detailed summary is presented in appendix 5 on page 203.

3.5.1.1 Utility Standards (US1-US7). These standards are meant to ensure that the information generated from evaluations serves the needs of intended users (Stufflebeam (2001, 2005), Rogers (1992), Weiss (1983), Patton (1978, 1986, and 1997) by paying attention to the following associated standards: stakeholder identification, evaluator credibility, information scope and selection, and values identification plus report clarity, timeliness, dissemination and impact (see also Appendix A, p.196).

3.5.1.2 Feasibility Standards (FS1-FS3): Defined by three standards, "the feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic and frugal" (Sanders, 1994, p.63; see also Appendix A, p.196).

3.5.1.3 Propriety standards (PS1-PS8): “The propriety standards according to Sanders (1994), are intended to ensure that an evaluation was conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.” (p.81). There are eight associated criteria that have to be complied with under this category (see also Appendix C, p.202).

3.5.1.4 Accuracy standards (AS1-AS12): According to Sanders **(**1994) this category of standards is based upon 12 key considerations, “accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about features that determine worth or merit of the program being evaluated.” (p.125) (see also Appendix D, p.205).

3.6 Considerations in the Application of PES.

Some authors, such as Stufflebeam and Shinkfield (2007) and Chatterji (2005), have expressed reservations regarding the universal applicability of the PES in contexts other than the USA and Canada. They argue for consideration, among other issues, with regard to political contexts in the application of the PES.

Australia’s standards, for example, are much more to do with ethical issues, whilst New Zealand places high importance on involving the Maori population in the evaluation of policies affecting them. These countries based their standards on the PES but adopted them significantly to emphasise with ethical issues.

Nevertheless, whilst the African Evaluation guidelines of 2002 of the African Evaluation Association (AfrEA) for evaluating programmes recognise, for example, the importance of political considerations, its guidelines have, generally, adopted similar standards as those outlined by the PES. The PES, therefore, provided a relevant set of guidelines to assess the merit and worth of evaluations. Patton (1997) affirms that the PES provides the profession’s standards and principles for evaluating evaluations.

Therefore, in the case of this study of Metaevaluation, the PES was used for assessing the quality of selected evaluations of the World Bank on poverty-reduction interventions in Uganda.

3.7 Conclusions

The methodology for evaluating the quality of four selected evaluations of poverty reduction interventions funded by the World Bank through the PRSCs was based on the mixed methods approach, grounded in constructivist evaluation within the context of the constructivist philosophical paradigm. The mixed methodological approach was composed of a case study approach, document analysis and descriptive frequency statistics. The selected mixed methods assisted in identifying evident causes of such remarkable inconsistencies in evaluation outcomes of selected evaluations and answering the specified research question.

In assisting to identify evident causes of such remarkable inconsistencies, selected mixed methods provided the appropriate leverage for triangulating multiple perspectives and information sources in order to generate credible evidence. Triangulation also insulated the study against the generation of evidence based on inappropriately construed pieces of evidence based on a single method. The use of mixed methods also protected the evaluation of the quality of evaluations from probable bias associated with the use of a single method to generate empirical evidence.

The philosophical orientation of the methodology was grounded in the constructivist paradigm and situated within the constructivist evaluation using the four selected evaluation cases funded through the PRSC by the World Bank in Uganda. On the basis of selected evaluation cases, a case study approach was appropriate in addressing accuracy issues by employing and triangulating multiple perspectives and information sources.

The closeness of selected evaluation cases to real–life situations provided a wealth of details and useful strands of knowledge. In addition, the use of multiple study cases enhanced the credibility of evidence and supported the manner in which the information was interpreted from the possible biases of external observers. In particular, the use of document analysis involved scrutinising, identifying, choosing, valuing and integrating data obtained from documents using the four selected evaluation cases. The prescribed quantitative scales based on the quality checklist prescribed under the PES were used for assessing the quality of evaluations. The judgments regarding the adequacy and strength of each selected evaluation case in meeting the requirements of a good evaluation stipulated under the PES were made based on a set of prescribed quantitative scores.

Each of the four selected evaluation cases in the Ugandan case study was subjected to a checklist of 300 requirements, composed of 1,200 checklists for all four selected evaluation cases. The four compulsory associated standards specified under the PES were: Service Orientation (P1), Valid Information (A5), Justified Conclusions (A10) or Impartial Reporting (A11) on any of the four compulsory associated standards. The judgments regarding the adequacy and strength of each selected evaluation case in meeting the requirements of a good evaluation stipulated under the PES were made as follows: 0-2 Poor, 3-4 Fair, 5-6 Good, 7-8 Very Good, 9-10 Excellent.

Descriptive frequency statistics were also used to summarise the data, which was generated from document analysis using SPSS based on the above prescribed quantitative scores stipulated under the PES. Summarised emerging frequency patterns supported the evaluation process in generating new insights into the behaviour of outcomes and assisted in creating meaningful interpretations of the results of the assessment based on emerging patterns.

CHAPTER 4: PRESENTATION OF THE FINDINGS

4.0 Introduction

The previous chapter outlined the methodology for evaluating the quality of the four selected evaluations of four poverty-reduction projects funded by the World Bank through its PRSCs in Uganda, based on the quality checklist stipulated in the PES.

Six metaevaluation conceptualisation premises outlined in the Methodology Chapter (section 3.2.3) were adopted as guidelines for the purposes of making sense in the analysis and interpretation of metaevaluation results. It is within the context of these metaevaluation conceptualisation premises that the application of the PES was situated.

The description of the findings on the four evaluations cases are presented in this chapter based on the prescribed metaevaluation checklist (summarised in Appendices A-D, p.196-205) under the PES so that possible causes of inconsistencies in the effectiveness of evaluation outcomes of poverty-reduction interventions by the World Bank identified in Chapter Two of this study could be clarified.

This chapter is divided into two major sections: the first presents metaevaluation findings on each evaluation case from document analysis using 300 checklists based on the following set of prescribed quantitative scales: 0-2 Poor, 3-4 Fair, 5-6 Good, 78 Very Good, 9-10 Excellent. Stufflebeam (1999) recommends that an evaluation be considered as failed according to the PES if it scores 'Poor' on Standard P1 Service Orientation, A5 Valid Information, A10 Justified conclusions or A11 Impartial Reporting, and that the judgments regarding the adequacy of the subject evaluation in meeting the standards can be made as follows: 0-2 Poor, 3-4 Fair, 5-6 Good, 7-8 Very Good, 9-10 Excellent". (p.2).

In order to identify descriptive frequency statistics patterns that emerged across the four selected evaluations and the checklist based on the PES across individual evaluation standards for all the four selected evaluation cases, the second major section presents cross-case descriptive frequency statistics findings based on each individual evaluation standard as well as associated criteria.

Summaries of overall findings from document analysis across all four selected evaluation cases are provided in Table F1 on p.225. In addition, a summary of findings from document analysis across these four selected evaluation cases, based on compulsory associated standards under the PES, are presented in Table G1 on p.227. Summaries of descriptive statistics based on all four selected evaluation cases and the four categories of evaluation standards under the PES are provided in Tables H1, H2, H3 and H4 on p.228.

4.1 Findings from Document Analysis based on Individual Evaluation Cases

4.1.1 The First Evaluation Case.

The findings from the examination of the strength of the first evaluation case, using the checklists provided under the PES, revealed the following patterns. These findings are also summarised generally in Table F1 on p.225 and specifically on compulsory associated standards under the PES in Table G1 on p.227.

4.1.1.1 Findings on Utility Standards (US1-US7).

The specific patterns of the findings that emerged for each of the checklist provided under the PES are described here below:

Stakeholder Identification (US1): The first evaluation case complied with two of ten requirements under this associated secondary standard. The results showed that the evaluation client was clearly identified and that potential stakeholders were consulted in order to identify their information needs. However, the results under this measure demonstrated that the evaluation case did not comply with eight requirements. There was no evidence in the evaluation case under consideration to show that leadership figures were engaged to identify other stakeholders, nor was there any information to show that stakeholders were used to identify other stakeholders. There was insufficient evidence in the evaluation case to provide substantial grounds to determine whether or not stakeholders were ranked for relative importance with the client. There were also not enough clues to show that any arrangement was made to involve stakeholders throughout the evaluation.

Based on the examination of the contents of the evaluation case, no evidence was found to confirm that the evaluation process was kept open to serve newly-identified stakeholders, and there were no indications in the evaluation case under review showing that the evaluation addressed the needs of the stakeholders. There was also no proof in the evaluation report to enable the researcher to judge whether or not both appropriate ranges of individual stakeholders and organisations were served. On this measure alone, the strength of the evaluation case was found to be poor.

Evaluator Credibility (US2): The first evaluation case complied with four out of ten requirements under this associated standard. The examination of the contents of this first evaluation case showed that competent evaluators were engaged and evaluators gave "stakeholders information on the evaluation plan's technical quality and practicality" (see Appendix A, p.196). In addition, the information reviewed confirms that evaluators stayed abreast of social and political forces.

The examination of the contents of the evaluation case demonstrated that evaluators kept some interested parties informed about the evaluation's progress. However, the evaluation failed to comply with six requirements. It is not clear in this evaluation case whether evaluators that the stakeholders trust were engaged. Neither was it clear from the examination of the evaluation case that evaluators engaged were appropriately responsive to gender issues, socio-economic status or race, language and cultural differences. There was also no evidence in the evaluation case, which revealed that the evaluation plans of selected interventions responded to key concerns of stakeholders, as well as whether or not stakeholders were helped to understand the evaluation plan. In addition, the assessment of the evaluation case did not show that evaluators attended appropriately to stakeholders' criticisms and suggestions. On this associated standard, the strength of this evaluation case was, therefore, adjudged to be fair.

Information Scope and Selection (US3): Five of the ten requirements under this measure were not met. The assessment of the nature of the selection and scope of information in the valuation case showed that the client's most important evaluation requirements were understood and that priority was assigned to the most important stakeholders. The examination of the contents of the evaluation case also revealed that priority was assigned to the most important questions. The assessment of the contents also indicated that information that was obtained was sufficient to address the stakeholders' most important evaluation question, and that evaluation effort was allocated in accordance with the priorities assigned to the information required.

Nevertheless the assessment of the evaluation case revealed that the evaluation case did not comply with the other five requirements. The review of the contents of the first evaluation case revealed that stakeholders were not interviewed to determine their different perspectives. There was no sign in the contents of the evaluation case that provided clues that there was assurance in any form, demonstrating that persons conducting the evaluation included the client in negotiating with a limited range of relevant interested parties, questions and information required. There was no evidence in the evaluation case which showed that flexibility was provided to include additional questions during the evaluation process. There was also no evidence to demonstrate whether or not sufficient information was obtained for assessing the merits of the programme. There were no indications in the contents of the evaluation case, which showed that sufficient information was obtained to assess the programme's worth. Therefore, due to compliance with five out of ten requirements, the strength of the evaluation case on this associated evaluation standard was found to be good.

Values Identification (US4): The first evaluation case complied with four out of the ten requirements under this measure. Analysis of the contents of the evaluation case revealed that a clear, defensible basis for value judgments was provided and that the evaluation determined the appropriate party(ies) to make the evaluation interpretations. In addition, the review of the contents indicated that reference was appropriately made regarding the relevant institutional mission and regarding the programme's goals.

Nevertheless, there were six requirements that were not met under this standard. The outcomes of the assessment revealed the following: alternative sources of values for interpreting evaluation findings were not considered; pertinent societal needs were not identified; pertinent customer needs were not identified; reference was not made to pertinent laws; stakeholders' values were not taken into account; and alternative interpretations were not appropriately presented, based on conflicting but credible values. The strength of this evaluation case was judged to be fair with a compliance level of four of the ten requirements on this associated standard.

Report Clarity (US5): On the basis of this measure, nine of the ten requirements were met. Examination of the contents of the evaluation case, guided by the prescriptions in the PES, confirms the following: essential information was clearly reported; the reports were brief, simple and direct; the reports focused on contracted questions; the programme and its contents were described; the purposes, procedures and findings of the evaluation were described; the conclusions and recommendations were supported; reporting in technical jargon was avoided, an executive summary was provided; and a technical report was provided. However, it was not evident in the assessment of the evaluation case that the evaluation was not reported in the language(s) of stakeholders. With a compliance level of nine requirements out of ten, the strength of the evaluation case was found to be excellent on this associated standard.

Report Timeliness and Dissemination (US6): Seven of the ten requirements under this measure were complied with in the first evaluation case. The assessment of this case showed that the interim reports were made on a timely basis to "intended users; the final report was delivered when it was needed; there were timely exchanges with the programme's 'policy board; there were timely exchanges with the program's staff; there were timely exchanges with the full range of right-to-know audiences; effective media were employed for reaching and informing the different audiences and the presentations were appropriately kept brief"(see also Table F1 on p.225).

However, the assessment of the evaluation case showed that the timely exchanges were not made with both the programme's customers and the public media, and that the audience was not helped to relate the findings to practical situations. The strength of the evaluation case was very good as it complied with seven requirements under this standard.

Evaluation impact (US7): On the basis of this measure, it was found from the assessment of the first evaluation case that out of ten requirements, four were met. Evidence from the evaluation case revealed that contact was maintained with the audience; stakeholders' use of the findings was encouraged and supported; interim reports were provided; and reports were open, frank and concrete. On the other hand, an examination of the contents of the first evaluation case revealed that stakeholders were not involved throughout the evaluation nor were they shown "how they might use the findings in their work" (see also results in Table F1 on p.225). In addition, the examination of the evaluation case showed that potential uses of findings were not forecast or addressed; written reports where not supplemented with ongoing oral communication.

There was no evidence in the evaluation reports, which suggested that feedback workshops were conducted on how to apply findings, and arrangements were not made to "provide follow-up assistance in interpreting and applying the findings" (see results in also Table F1 on p.225). With a compliance level of four requirements out of ten, the strength of the evaluation was found to be fair. In summary, with three fair scores, and one each poor, good, very good and excellent score under the utility standard, overall, the strength of the evaluation case on this standard is predominantly fair.

4.1.1.2 Findings on Feasibility Standards (FS1-FS3).

The findings from the assessment of the extent to which the first evaluation case complied with the requirements of the PES under each of the associated standards are presented below.

Practical Procedures (FS1): The first evaluation case complied with eight of the ten requirements under this measure. Evidence generated from the assessment showed that the evaluation under consideration tailored the methods and instruments to information requirements; disruption were minimised; the data burden was minimised, competent staff were appointed; staff were trained; the evaluation process chose procedures that the staff were qualified to carry out; procedures in light of known constraints were chosen and a realistic schedule was made. However, there is no evidence in the valuation case that showed that locals were engaged to help conduct the evaluation. It is also not clear from the evaluation case that evaluation procedures were appropriately made as part of routine events. The evaluation complied with eight requirements out of ten; therefore, its strength was judged to be very good.

Political Viability (FS2): The first evaluation case complied with eight of the ten requirements under this measure. Evidence generated from the assessment shows that cooperation was fostered; there was agreement on the editorial and dissemination authority; interim reports were issued and a firm public contract was employed. However, the results of the assessment also revealed that stakeholders were not involved throughout the evaluation; divergent views were not reported; it is not clear whether or not all right-to-know audiences received the report nor whether any corrupted evaluation was terminated. The review of the evaluation case did not indicate whether or not differing positions of different interest groups were anticipated and finally, no evidence in the evaluation case suggested that attempts of bias or misapplication of the findings were averted or counteracted. The evaluation case complied with four requirements out of ten. Therefore, the strength of the evaluation case on this measure was found to be fair.

Cost-effectiveness (FS3): The first evaluation case complied with seven of the ten requirements under this measure. The assessment of the evaluation report showed that the evaluation was: efficient; produced information worth the investment; informed decisions; fostered programme improvement; provided accountability information; generated new insights and minimised disruptions. Nevertheless, there was no evidence in the evaluation case that suggested that the evaluation process made use of in-kind services, helped spread effective practices nor minimised time demands on programme personnel. The evaluation case under consideration complied with seven requirements out of ten and the strength of the case on this measure was found to be very good. With two scores of very good under the feasibility standard, the overall strength of the evaluation case on this standard was found to be good.

4.1.1.3 Findings on Propriety Standards (PS1-PS8).

The results from the assessment of the extent to which the first evaluation case complied with these propriety standards are described in the following sections.

Service Orientation (PS1): The examination of the evaluation case revealed that the evaluation: assessed the needs of the programme's customers; promoted excellent service; made the evaluation's service orientation clear to stakeholders; gave interim feedback for programme improvement; identified programme strengths on which to build and identified programme weaknesses to correct. By contrast, there is no evidence in the evaluation case, which suggested that the evaluation: helped to ensure that the full range of rightful programme beneficiaries was served; that programme outcomes were assessed against target customers' assessed needs; that harmful practices were exposed and that all right-to-know audiences were informed of the programme's positive and negative outcomes. On this measure alone, the evaluation complied with six requirements out of ten. The strength of the evaluation was, therefore, judged to be good.

Formal Agreements (PS2): In seeking to determine whether or not the evaluation complied with the formal agreements, scrutiny of the evaluation case reveals that: formal parties agreed on the evaluation purpose and questions, audiences, evaluation reports, editing, release of reports, evaluation procedures and schedule, evaluation staff and evaluation resources. Inspection of the degree of compliance with the PES revealed that formal parties did not agree on both the confidentiality/anonymity of data and

Metaevaluation. The strength of the evaluation was found to be very good because it complied with eight out of ten requirements.

Rights of Human Subjects (PS3): The evidence generated from the scrutiny of the evaluation case under consideration demonstrated that the process of designing and conducting the evaluation made clear to stakeholders that the evaluation respected and protected the rights of human subjects. It is also clear in the evaluation case that the intended uses of the evaluation were clarified, and that the evaluation process followed due process. Furthermore it is clear from the evaluation that participant values were understood, that protocol was followed and that confidentiality/anonymity agreements were honoured.

The examination of the evaluation case, nevertheless, showed that the designing and conducting of the evaluation did not keep stakeholders informed. Insights into the contents of the evaluation case also revealed that the evaluation process did not uphold civil rights, did not respect diversity and, based on the contents of the evaluation, it was not clear whether or not the evaluation did any harm. On this measure, the evaluation complied with six requirements out of ten and the overall strength of the evaluation was found to be good.

Human Interactions (PS4): On this measure, the results showed that evaluators consistently related to all stakeholders in a professional manner; maintained effective communication with stakeholders; followed the institution's protocol; minimized disruption and honoured time commitments. There was, however, no evidence in the evaluation case to suggest that evaluators were alerted to, and addressed participants' concerns regarding the evaluation.

There is insufficient information to conclude whether or not the evaluation process was sensitive to participants' diversity of values and cultural differences, and if the evaluation process was even-handed in addressing different stakeholders. It is not clear, based on the contents of the evaluation case, whether it can be stated that the evaluation process neither ignored nor helped to cover up any participant's incompetence, unethical behaviour, fraud, waste or abuse, or if the evaluation process honoured participants' privacy rights. The evaluation on this measure was judged to be good based on its compliance with five requirements out of ten.

Complete and Fair Assessment (PS5): The evidence generated demonstrates that the evaluation case complied with five out of ten requirements. The results showed that: the evaluation case assessed and reported the programme's strengths; assessed and reported its weaknesses; reported on intended outcomes; gave a thorough account of the evaluation's process and had the draft report reviewed. Nevertheless, evidence drawn from the contents of the evaluation case shows that the evaluation process: did not report on unintended outcomes; did not, as appropriate, show how the programme's strengths could be used to overcome its weaknesses, did not appropriately address criticisms of the draft report; did not acknowledge the final report's limitations and did not estimate and report the effects of the evaluation's limitations on the overall judgment of the programme. The strength of the evaluation was found to be good because it complied with five out ten requirements.

Disclosure of Findings (PS6): On this measure the evaluation complied with five requirements out of ten and the strength of the evaluation was adjudged to be good. In this case, results indicate that: the formal parties defined the right-to-know audiences; informed the audiences of the evaluation's purposes and projected reports; reported all findings in writing; reported balanced, informed conclusions and recommendations; and showed the basis for the conclusions and recommendations. However, the evaluation did not establish a contractual basis for complying with right to-know requirements nor did it report relevant points of view of both the supporters and the critics of the programme. In addition, the evaluation process did not disclose the evaluation's limitations in reporting; did not adhere strictly to a code of directness, openness and completeness; and did not ensure that reports reach their audiences.

Conflict of Interest (PS7): On this measure, the results showed that in dealing with "conflict of interest", the evaluation process engaged multiple evaluators; maintained evaluation records for independent review; contracted with the funding authority rather than the funded programme; and had internal evaluators report directly to the chief executive officer. However, the assessment of the evaluation case showed that the evaluation process did not identify potential conflicts of interest early in the evaluation; provide written, contractual safeguards against identified conflicts of interest; did not engage independent parties, as appropriate, to assess the evaluation for its susceptibility or corruption by conflicts of interest and did not release evaluation procedures. The assessment of the evaluation case also showed that data, and reports for public review when appropriate; did not report equitably to all right-to-know audiences; and did not engage uniquely-qualified persons to participate in the evaluation even if they had a potential conflict of interest nor did it take steps to counteract the conflict. On this measure, the evaluation complied with four requirements and the strength of the evaluation was judged to be fair.

Fiscal Responsibility (PS8): On this case, results indicate that the evaluation specified and budgeted for expense items in advance; obtained appropriate approval for necessary budgetary modifications; assigned responsibility for managing the evaluation finances; maintained accurate records of sources of funding and expenditure; maintained adequate personnel records concerning job allocations and time spent on the job; employed comparison shopping for evaluation materials and was frugal in expending evaluation resources. There was no evidence provided in the evaluation case, which indicated that budgets were kept sufficiently flexible to permit appropriate re-allocations to strengthen the evaluation and whether the valuation process employed comparison contract bidding. On this measure, the evaluation complied with seven requirements out of ten and the strength of the evaluation was judged to be very good. With two scores of good under propriety standards, the overall strength of the evaluation case on this standard was found to be good.

4.1.1.4 Findings on Accuracy Standards (AS1-AS12).

The results from the assessment of the extent to which the first evaluation case complied with the requirements under each of these associated standards are discussed below.

Programme Documentation (AS1): The evidence generated from assessing the evaluation demonstrates that: descriptions of the intended programme from various written sources were collected; descriptions of the intended programme from the client and various stakeholders were collected; the manner in which the programme was intended to function was described; records from various sources of how the programme operated were maintained; the evaluation described how the programme actually functioned; in addition, the evaluation analysed: discrepancies between the various descriptions of how the programme was intended to function; and discrepancies between how the programme was intended to operate and how it actually operated. The assessment also shows that the evaluation produced a technical report that documented the programme's operations.

However, there was no evidence in the evaluation to indicate that the evaluation feasibly engaged independent observers to describe the programme's actual operations. There was also no evidence to suggest that the client and various stakeholders were asked to assess the accuracy of recorded descriptions of either the intended, or the actual, programme. On this measure, the evaluation complied with eight requirements out of ten and the strength of the evaluation was judged to be very good.

Context Analysis (AS2): An examination of the evaluation case in this context indicates that multiple sources of information were used to describe the programme's context; that the context's technical, social, political, organisational and economic features were described; and that the evaluation analysed how the programme's context was similar to, or different from, contexts where the programme might had to be adopted.

However, there was no evidence in the report, which indicated that a log of unusual circumstances was maintained; that the evaluation recorded instances in which individuals or groups intentionally or otherwise interfered with the programme; that the evaluation recorded instances in which individuals or groups intentionally or otherwise gave special assistance to the programme; that the evaluation reported those contextual influences that appeared to influence the programme significantly and which might have been of interest to potential adopters; that the evaluation estimated the effects of context on programme outcomes; that the evaluation identified and described any critical competitors to this programme that functioned at the same time and in the programme's environment; that the evaluation described how people in the programme's general area perceived its existence, importance and quality. With compliance on three requirements out of ten, the strength of the evaluation case on this standard was found to be fair.

Description of Purposes and Procedures (AS3): The evidence from the assessment shows that at the evaluation's outset, the client's purposes for the evaluation were recorded; actual evaluation procedures, as implemented, were recorded; and the evaluation's purposes and procedures in the summary and full length evaluation reports were described. However, there was no evidence to enable the researcher to make plausible judgment as to whether or not stakeholders' intended uses of evaluation findings were described and monitored.

No sufficient evidence was found in the evaluation, which suggested how the evaluation's purposes were described and monitored, stayed the same or changed over time; neither were there found any points of agreement and disagreement among stakeholders regarding the evaluation's purposes identified and assessed. Furthermore, there was no evidence in the evaluation case, which showed whether or not, as appropriate, evaluation procedures to accommodate changes in the evaluation's purposes were updated; there was no indication in the evaluation case, which indicated that when interpreting findings, different stakeholders' intended uses of the evaluation were taken into account; also, it was not indicated that, when interpreting findings, the extent to which the intended procedures were effectively executed was also taken into account. The researcher could not ascertain, based on the assessment of the contents of the evaluation, that independent evaluators to monitor and evaluate the evaluation's purposes and procedures were feasibly engaged. On this measure, the evaluation met three requirements out ten requirements and the strength of the evaluation was found to be fair.

Defensible Information Sources (AS4): On this measure, the results show that the evaluation complied with two requirements out of ten. The evaluation shows that varieties of data collection methods were appropriately used and included data collection instruments in a technical appendix to the evaluation report. However, the results of the assessment of the evaluation could not provide sufficient information to determine whether or not the information was obtained from a variety of sources; previously collected information once validated was pertinent; information sources were documented and reported; the criteria and methods used to select information sources were documented, justified and reported.

There was also insufficient information to determine whether or not the population was defined for each source; whether or not any employed sample for each population was defined appropriately; whether or not the means used to obtain information from each source were documented, justified and reported; and whether or not any biasing features in the obtained information were documented and reported. The evaluation report complied with two requirements for this standard, therefore the strength of the evaluation was judged to be poor.

Valid Information (AS5): On the basis of this measure, evidence suggested that the evaluation process focused the evaluation activities on key questions, appropriately employed multiple measures to address each question, and trained and calibrated data collectors. However, there was insufficient information to judge whether or not detailed description of the constructs and behaviours was provided about which information was acquired; no evidence to judge whether the type of information each employed procedure acquired was assessed or reported. There was also insufficient information to judge whether or not the collection conditions and process were documented and reported; whether or not how information from each procedure was scored, analysed and interpreted was documented. Furthermore, there was insufficient information to judge whether or not inferences singly and in combination were reported and justified; whether or not the comprehensiveness of the information provided by the procedures as a set, in relation to the information needed to answer the set of evaluation questions, was assessed and reported; and whether or not meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures were established. On this standard, three requirements out of ten were met. The strength of the evaluation was, therefore, fair.

Reliable Information (AS6): The evaluation report complied with four requirements under this associated criteria. Evidence shows that the unit of analysis for each employed data collection device was specified; as feasible, measuring devices that in the past had shown acceptable levels of reliability for their intended uses were chosen, scorers and analysts to produce consistent results were trained and calibrated; and the consistency between multiple observers were appropriately checked and engaged.

However, the evaluation case did not comply with six requirements. The results show that the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions and the evaluator's bias in the reporting reliability of an instrument, were not assessed and reported; nor was the consistency of scoring, categorisation and coding checked or reported. In order to identify and control sources of error, pilot testing of new instruments was not conducted; reliability problems in the final report were not acknowledged; the effects of unreliability in the data on the overall judgment of the programme were not estimated or reported. With a compliance level of four requirements out of ten, the strength of the evaluation was fair.

Systematic Information (AS7): The results of the assessment indicate that protocols for quality control of the evaluation information were established; the evaluation staff adhered to the data protocols and verified data entry; staff were trained; data tables generated from computer output or other means were proofread and verified; storage of the evaluation information was systematised and controlled, and data providers verified the data they submitted. There was no evidence in the valuation report, which suggested that the accuracy of scoring and coding was systematically checked; multiple evaluators and checking the consistency of their work were used when feasible; those with access to the evaluation information were not defined; access to the evaluation information according to established protocols was not strictly controlled. On this measure, the evaluation complied with six requirements out of ten. In this case the strength of the evaluation was found to be good.

Analysis of Quantitative Information (AS8): The results showed that preliminary exploratory analyses were conducted to ensure the data's correctness and to gain a greater understanding of the type of data appropriate for the evaluation questions and nature of the data collection procedures which were chosen. The results also show that appropriate data was specified for each procedure and how its key assumptions were to be met. In some instances, visual displays were used to clarify the presentation and interpretation of statistical results.

This particular evaluation case, however, did not comply with six requirements. There was no evidence to suggest that limitations of each analytic procedure, including failure to meet assumptions, were reported; multiple analytic procedures to check on consistency and replicability of findings were employed; variability as well as central tendencies were examined. There was also no evidence to suggest that outliers and their correctness were identified, verified and examined; and whether or not statistical significance were assessed. The strength of this evaluation was found to be fair based on the four out of ten requirements that were met.

Analysis of Qualitative Information (AS9): Evidence showed that the evaluation's focus on key questions defined the boundaries of information to be used, chose analytic procedures and methods of summarisation that were appropriate to the evaluation questions, and employed qualitative information and derived a set of categories that is sufficient to document, illuminate and respond to the evaluation questions. However, the examination of the evaluation case showed that there was also no evidence based on the contents of the evaluation case which demonstrated that information keyed to the important evaluation questions was obtained; no evidence in the evaluation case showed that the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders was verified; no evidence in the evaluation case showed that the derived categories for reliability and validity were tested and no evidence in the evaluation case showed that the obtained information into the validated analysis categories were classified. While conclusions and recommendations were derived and demonstrated their meaningfulness; limitations of the referenced information, analyses, and inferences were not reported. The strength of the evaluation was judged to be fair based on a compliance level of four requirements out of ten requirements.

Justified Conclusions (AS10): The results showed that the evaluation focused conclusions directly on the evaluation questions; the evaluation procedures and findings were accurately reflected but the applicable time periods, contexts, purposes and activities were limited to conclusions. Furthermore, there was no information found in the contents of the evaluation case that suggested that information that supports each conclusion was cited; the programme's side effects were identified and reported; plausible alternative explanations of the findings were reported; there was no lead to explain why rival explanations were rejected nor was there any warning against making common misinterpretations. There was also no evidence in the evaluation case showed that the results of a prerelease review of the draft evaluation report were obtained and addressed the evaluation's limitations which were reported. Here the evaluation complied with three requirements out ten and the strength of the evaluation was found to be fair.

Impartial Reporting (AS11): The results showed that appropriate editorial authority was established, that the right-to-know audiences were determined and that appropriate plans for releasing findings to all right-to-know audiences were established and followed.

However, the contents of the evaluation case did not provide sufficient information to suggest that a wide range of clients were engaged in determining steps to ensure fair, impartial reports; reports were safeguarded from deliberate or inadvertent distortions; perspectives of all stakeholder groups were reported, alternative plausible conclusions were also reported; outside audits of reports were obtained; steps taken to control bias were described; and public presentations of the findings to help guard against and correct distortions by other interested parties were conducted. On this standard, the evaluation complied with three requirements out of ten. Therefore, the strength of this evaluation was found to be fair.

Metaevaluation (AS12): The results under this standard showed that responsibility was assigned to someone for documenting and assessing the evaluation process and products. However, sufficient information was not obtained in the evaluation case, which provides substantial grounding to demonstrate that the standards to be used in judging the evaluation were designated or defined; both formative and summative Metaevaluation were employed; the budget was appropriate and sufficient for conducting the Metaevaluation.

In addition, there was no evidence suggesting that a full range of information, needed to judge the evaluation against the stipulated standards, was recorded. An independent Metaevaluation was not feasibly contracted; audiences that would receive the Metaevaluation report were not determined and recorded; the instrumentation, data collection, data handling, coding and analysis against the relevant standards were not evaluated; the evaluation's involvement and communication of findings to stakeholders against the relevant standards were not evaluated; and a record of all Metaevaluation steps, information and analyses was not maintained. With the compliance of one out of ten requirements, the strength of the evaluation of this standard was judged to be poor, and the overall strength of the evaluation case study on the accuracy standard, judged to be fair.

4.1.2 The Second Evaluation Case Study.

4.1.2.1 Findings on Utility Standards (US1-US7).

The findings that emerged from the assessment, based on the PES for the second evaluation case, are described below.

Stakeholder Identification (US1): Under this measure, six out of ten requirements were met. The results showed that the evaluation client was clearly identified; engaged with leadership figures to identify other stakeholders; used stakeholders to identify other stakeholders; kept the evaluation open to serve newly identified stakeholders; addressed stakeholders' evaluation needs and served an appropriate range of individual stakeholders.

There was no evidence, however, in this second evaluation case under consideration to show that potential stakeholders were consulted in order to identify their information needs. There was no evidence in this case, which suggested that, together with the client, stakeholders were ranked for relative importance. There was also no evidence in the evaluation case, which either suggested that arrangements were made to involve stakeholders throughout the evaluation, or that the evaluation served an appropriate range of stakeholder organisations. Six requirements were met under this standard, and the strength of the evaluation case was found to be good.

Evaluator Credibility (US2): The assessment based on this standard showed that eight out of ten requirements were met under this associated standard. The results showed that: competent evaluators were engaged; that the stakeholders trust were engaged on a limited basis; evaluators that could address stakeholders’ concerns were engaged; evaluators that are appropriately responsive to issues of gender, socio-economic status, race, and language and cultural differences were engaged. There was also evidence assuring that the evaluation plan responded to a certain limited range of key stakeholders' concerns: such a limited range of stakeholders were helped to understand the evaluation plan; the evaluators stayed abreast of social and political forces; and interested parties were informed about the evaluation's progress.

However, there was no evidence in the evaluation report, which suggested that such a limited range of stakeholders were given information on the evaluation plan's technical quality and practicality, nor that the criticisms and suggestions of stakeholders were attended to appropriately. The strength of this evaluation case was judged to be very good with a compliance level of eight out of ten requirements on this associated standard.

Information Scope and Selection (US3): There were four requirements out of requirements that were met under this measure. Scrutiny of the selection and scope of information showed that the client's most important evaluation requirements were understood; that priority was assigned to which was deemed as the most important stakeholders; that priority was assigned to the most important questions; and that evaluation effort was allocated in accordance with the priorities assigned to the required information.

However, the evaluation did not comply with the remaining six requirements. The results of the assessment showed that stakeholders were not interviewed to determine their different perspectives. There was no evidence in the evaluation case that showed assurance that the evaluator and client negotiated pertinent audiences, questions and required information. There was no evidence in the evaluation case that illustrated that flexibility was allowed for adding questions during the evaluation process and finally, there was no evidence in the evaluation case that showed that sufficient information was obtained to address what was deemed as key stakeholders' most important evaluation questions. There was also substantial evidence to suggest that sufficient information was obtained to assess the programme's merit. Finally, an assessment of the evaluation case revealed that insufficient information was obtained to assess the programme's worth. Therefore, with compliance of four requirements out ten requirements, the strength of the evaluation case on this associated standard was found to be fair.

Values Identification (US4): The second evaluation case complied with seven out of ten requirements under this measure. Evidence showed that the evaluation under consideration provided a clear, defensible basis for value judgments; that reference was appropriately made regarding the relevant institutional mission; reference was made regarding the goals of the programme; pertinent societal needs were identified; pertinent customer needs were identified; reference was made to pertinent laws; stakeholders' values were taken into account; and alternative interpretations were appropriately presented based on conflicting but credible values.

There were three requirements that were not met under this associated standard. The outcomes of the assessment showed that: appropriate party(ies) to make the evaluational interpretations' alternative sources of values for interpreting evaluation findings were not considered; the stakeholders' values were not taken into account; and alternative sources of values for interpreting evaluation findings were not considered. The strength of this evaluation case was judged to be very good with a compliance level of seven out of ten requirements on this associated standard.

Report Clarity (US5): On the basis of this measure, all ten requirements were met. Examination of the contents of the evaluation case, guided by the prescriptions in the PES, confirm the following: essential information was clearly reported; the reports were brief, simple and direct; the reports focused on contracted questions; the programme and its contents were described; the purposes, procedures and findings of the evaluation were described; conclusions and recommendations were supported; reporting in technical jargon was avoided; an executive summary was provided; the evaluation was reported in the language(s) of stakeholders; and a technical report was provided. With a compliance level of ten requirements out of ten, the strength of the evaluation cases was found to be excellent on this associated standard.

Report Timeliness and Dissemination (US6): The second evaluation case complied with seven out of ten requirements under this measure. The assessment of the reports showed that the interim reports were made on a timely basis to intended users; the final report was delivered when it was needed; there were timely exchanges with the programme's policy board; there were timely exchanges with the programme's staff; there were timely exchanges with the programme's customers and the presentations were kept appropriately brief. The contents of the evaluation cases showed that examples were used to help audiences to relate the findings to practical situations.

However, based on the contents of the evaluation, there was no evidence to suggest that there were timely exchanges with the full range of right-to-know audiences; that there were timely exchanges made with both the programme's customers and the public media; neither was effective media for reaching and informing the different audiences employed. The strength of the evaluation case was very good as it complied with seven out of ten requirements under this associated standard.

Evaluation Impact (US7): On the basis of this measure, it was found from the assessment of the report that seven out of ten requirements were met. The findings revealed that a limited range of stakeholders were involved throughout the evaluation; such a limited range of stakeholders were shown how they might use the findings in their work; potential uses of findings were forecast and addressed; and interim reports were provided. Evidence in the evaluation reports also suggested that the evaluation report was open, frank and concrete; arrangements were made to provide follow-up assistance in interpreting and applying the findings. On the other hand, there was no evidence in the second evaluation case, which suggested that contact was maintained with such a limited range of stakeholders during the evaluation process; that written reports were supplemented with ongoing oral communication, or that feedback workshops were conducted to go over and apply findings. With a compliance level of seven requirements out of ten, the strength of the evaluation was found to be very good.

In summary, with four very good scores, one fair score, one good score and one excellent score under the utility standard, overall the strength of the evaluation case on this standard is good.

4.1.2.2 Findings on Feasibility Standards (FS1-FS3).

The specific patterns of the findings for this evaluation standard are described below.

Practical Procedures (FS1): The second evaluation case complied with eight out of ten requirements under this measure. Evidence generated from the assessment showed that the evaluation process for this case tailored the methods and instruments to information requirements; minimised the data burden; appointed competent staff; trained staff; chose procedures that the staff were qualified to carry out; chose procedures in light of known constraints and that a realistic schedule was made.

However, there was no evidence showing that disruptions were minimised, nor that local parties of interest in the poverty reduction intervention were engaged to help conduct the evaluation. It is not clear from the evaluation that evaluation procedures were appropriately made part of routine events. The evaluation complied with seven requirements out of ten. Therefore, its strength was judged to be very good

Political Viability (FS2). The second evaluation case complied with eight out of ten requirements under this measure. Evidence generated from the assessment showed that cooperation was fostered; there was agreement on the editorial and dissemination authority; stakeholders were involved throughout the evaluation; interim reports were issued; divergent views were reported and a firm public contract was employed. However, it is not clear whether all right-to-know audiences received the report; it is also not clear whether any corrupted evaluation was terminated. Furthermore, the examination of the contents of the evaluation case did not indicate whether differing positions of different interest groups were anticipated and, finally, there was no evidence to suggest that attempts of bias or misapplication of the findings were averted or counteracted. The evaluation case complied with six requirements out of ten. Therefore, the strength of the evaluation case on this measure was found to be good.

Cost-effectiveness (FS3): The second evaluation case complied with seven out of ten requirements under this measure. The assessment of the evaluation report showed that the evaluation was: efficient; informed decisions; fostered programme improvement; provided accountability information; generated new insights; helped to spread effective practices; and minimised disruptions. However, there was no evidence in the evaluation case to suggest that the evaluation produced information worth the investment, nor minimised both disruption and time demands on programme personnel. The evaluation case under consideration complied with seven requirements out of ten and the strength of the case on this measure was found to be very good. With two scores of good under the feasibility standard, the overall strength of the evaluation case on this standard was found to be very good.

4.1.2.3 Findings on Propriety Standards (PS1-PS8).

The specific patterns of the findings that emerged for each of the above associated evaluation standard are described below.

Service Orientation (PS1): The results showed that: the evaluation assessed the needs of the programme’s customers; assessed programme outcomes against targeted customers’ assessed needs; helped to ensure that the full range of rightful programme beneficiaries be served; promoted excellent service; made the evaluation’s service orientation clear to stakeholders; identified programme strengths on which to build, identified programme weaknesses that required correction; and informed all right-to-know audiences of the programme's positive and negative outcomes. By contrast, there is no evidence in the evaluation case to suggest that interim feedback for programme improvement nor exposed harmful practices were provided to a wide range of stakeholders. On this measure alone, the evaluation complied with eight requirements out of ten. The strength of the evaluation was, therefore, judged to be very good.

Formal Agreements (PS2): The results showed that formal parties agreed on the evaluation purpose and questions, audiences, evaluation reports, editing, release of reports, evaluation staff, and evaluation resources. The results of the assessment, however, revealed that formal parties did not agree on the confidentiality/anonymity of data, evaluation procedures and schedule, or Metaevaluation. The strength of the evaluation was found to be very good because it complied with seven out ten requirements.

Rights of Human Subjects (PS3): The evidence generated demonstrated that the process of designing and conducting the evaluation clarified intended uses of the evaluation; followed due process; upheld civil rights; understood participant values; followed protocol; honoured confidentiality/anonymity agreements and did no harm. The examination of the evaluation case showed that the designing and conducting of the evaluation did not, however, make clear to a wide range of stakeholders that the evaluation would respect and protect the rights of human subjects, keep stakeholders informed and respect diversity. On this measure, the evaluation complied with seven requirements and the overall strength of the evaluation was found to be very good

Human Interactions (PS4): On this measure, the results showed that evaluators consistently related to all stakeholders in a professional manner; maintained effective communication with stakeholders; followed the institution's protocol; minimised disruption; honoured participants' privacy rights and honoured time commitments. There is also no evidence in the report, which suggested that the evaluation ignored or helped to cover up any participant's incompetence, unethical behaviour, fraud, waste or abuse.

However, no evidence in the evaluation case suggested that evaluators were alert to, or addressed participants' concerns regarding the evaluation; were sensitive to participants' diversity of values and cultural differences or were even-handed in addressing different stakeholders. The evaluation on this measure was judged to be very good, based on its compliance with seven requirements out of ten.

Complete and Fair Assessment (PS5): The evidence generated demonstrated that the evaluation case complied with six out of ten requirements. The results showed that the evaluation case assessed and reported the programme's strengths; assessed and reported the programme's weaknesses; reported on intended outcomes; reported on unintended outcomes; had the draft report reviewed and acknowledged the final report's limitations.

Nevertheless, the evaluation did not give a thorough account of its process, as appropriate, and did not show how the programme's strengths could be used to overcome its weaknesses. The effects of the evaluation's limitations on the overall judgment of the programme were neither estimated nor reported. In addition, there was no evidence, which suggested that criticisms of the draft report were appropriately addressed. On this associated standard, the strength of the evaluation was found to be good because it complied with six out ten requirements.

Disclosure of Findings (PS6): On this measure the evaluation complied with four requirements out of ten and the strength of the evaluation was judged to be fair. In this case, results indicated that the formal parties defined the right-to-know audiences; reported all findings in writing; showed the basis for the conclusions and recommendations; and in reporting, adhered strictly to a code of directness, openness and completeness. However, the evaluation did not establish a contractual basis for complying with right-to-know requirements; did not report relevant points of view of both supporters and critics of the programme; did not disclose the evaluation's limitations; in reporting, did not adhere strictly to a code of directness, openness and completeness; and did not ensure that reports reached their intended audiences.

Conflict of Interest (PS7): In the case of the second evaluation, the results showed that in dealing with conflict of interest, the evaluation process engaged multiple evaluators; maintained evaluation records for independent review; and when appropriate, released evaluation procedures, data and reports for public review and contracting with the funding authority rather than the funded programme.

However, the evidence, based on the contents of the evaluation case, showed that in dealing with conflict of interest, the evaluation process did not: identify potential conflicts of interest early in the evaluation; provide written or contractual safeguards against identified conflicts of interest; did not, as appropriate, engage independent parties to assess the evaluation for its susceptibility to corruption by conflicts of interest; did not have internal evaluators report directly to the chief executive officer; report equitably to all right-to-know audiences; engage uniquely-qualified persons to participate in the evaluation, even if they had a potential conflict of interest; and could not take steps to counteract the conflict. On this measure, the evaluation complied with four requirements out of ten and the strength of the evaluation was judged to be fair.

Fiscal Responsibility (PS8): The results under this standard indicated that the evaluation specified, and budgeted for expense, items in advance; kept the budget sufficiently flexible to permit appropriate reallocations to strengthen the evaluation; obtained appropriate approval for necessary budgetary modifications; and assigned responsibility for managing the evaluation finances, and evidence suggested that the evaluation process was frugal in expending evaluation resources.

There was no evidence provided in the evaluation case, which indicated that accurate records of sources of funding and expenditure were maintained; that comparison shopping for evaluation materials were employed; that comparison contract bidding was employed; or that, as appropriate, an expenditure summary as part of the public evaluation report was included. On this measure, the evaluation complied with five requirements out of ten and the strength of the evaluation was judged to be good. In summary, the overall strength of the evaluation case on this standard was found to be good.

4.1.2.4 Findings on Accuracy Standards (AS1-AS12).

The specific patterns of the findings that emerged for each of the above associated evaluation standard are described below.

Programme Documentation (AS1): The evidence generated from assessing the evaluation demonstrated that descriptions of the intended programme from various written sources were collected; the manner in which the programme was intended to function was described; records from various sources of how the programme operated were maintained; the manner in which the programme actually functioned was described; discrepancies between the various descriptions of the way in which the programme was intended to function were analysed; discrepancies between the way in which the programme was intended to operate and the way it actually operated were analysed; and a technical report documenting the programme's operations was produced.

However, there was no evidence in the evaluation case to indicate that descriptions of the intended programme from the client and various stakeholders were collected, as feasible; that independent observers to describe the programme's actual operations were engaged; and, whether or not both the client and the various stakeholders were asked to assess the accuracy of recorded descriptions of both the intended and the actual programme. On this measure, the evaluation complied with seven requirements out of ten and the strength of the evaluation was judged to be very good.

Context Analysis (AS2): An examination of the evaluation case indicated that multiple sources of information to describe the programme's context were used; the context's technical, social, political, organisational and economic features were described; a log of unusual circumstances was maintained; the way in which the programme's context is similar to, or different from, contexts where the programme might have been adapted was analysed; and the effects of context on programme outcomes were estimated.

However, there was no evidence in the report, which indicated that instances in which individuals or groups, intentionally or otherwise, interfered with the programme were recorded; instances in which individuals or groups, intentionally or otherwise, gave special assistance to the programme were recorded; those contextual influences that appeared to influence the programme significantly and that might be of interest to potential adopters were reported: any critical competitors to this programme that functioned at the same time and in the programme's environment were identified and described; and, the way in which people in the programme's general area perceived the programme's existence, importance and quality were described. With compliance of five requirements out of ten, the strength of the evaluation case on this standard was found to be good.

Describe Purposes and Procedures (AS3): The evidence from the assessment showed that at the evaluation's outset, the client's purposes for the evaluation were recorded; stakeholders' intended uses of evaluation findings were monitored and described; and the actual evaluation procedures, as implemented, were recorded.

However, there was no evidence to enable the researcher to make a plausible judgment as to whether or not the evaluation's purposes were monitored and described to determine whether or not they stayed the same or changed over time; points of agreement and disagreement among stakeholders regarding the valuation's purposes were identified and assessed; evaluation procedures to accommodate changes in the evaluation's purposes were, as appropriate, updated; the different stakeholders' intended uses of the evaluation, when interpreting findings, were taken into account. Furthermore, the extent to which the intended procedures were effectively executed was taken into account when interpreting findings; the evaluation's purposes and procedures in the summary and full-length evaluation reports were described; and, it was not clear as to whether or not independent evaluators to monitor and evaluate the evaluation's purposes and procedures were, as feasible, engaged. On this measure, the evaluation, met three out of ten requirements and the strength of the evaluation was found to be fair.

Defensible Information Sources (AS4): On this measure, the results showed that the evaluation complied with six requirements out of ten. The examination of the evaluation case showed that information from a variety of sources was obtained, as appropriate: a variety of data collection methods were employed: report information sources were documented: the criteria and methods used to select information sources were documented, justified and reported; data collection instruments were included in the appendix to the evaluation report; and any biasing features in the obtained information were documented and reported.

However, the assessment of the evaluation could not provide sufficient information to determine whether or not the: previously collected information, once validated, was used pertinently; for each source, the population was defined; any employed sample was defined; and the means used to obtain information from each source were documented, justified and reported. The evaluation report complied with six requirements for this standard, therefore, the strength of the evaluation was judged to be good.

Valid Information (AS5): On the basis of this measure, evidence suggested that the process of evaluation focused on key questions; provided a detailed description of the constructs and behaviours, about which information would be acquired; trained and calibrated the data collectors; documented the way in which information from each procedure was scored, analysed and interpreted; assessed and reported the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions; and, whether or not the process established meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures.

However, there was insufficient information to enable the researcher to judge whether or not: multiple measures to address each question were employed appropriately in order to assess and report on the type of information each employed procedure acquired; data collection conditions and processes were documented and reported; and, finally, inferences, singly and in combination, were reported and justified. On this standard, three requirements out of ten were met. The strength of the evaluation was, therefore, good.

Reliable Information (AS6): The evaluation report complied with two requirements under this associated criteria. The results showed that for each employed data collection device, the unit of analysis was specified and reliability problems in the final report were acknowledged.

However, the evaluation case did not comply with the following eight requirements: There was no evidence in the evaluation case to ascertain that preliminary exploratory analyses were conducted to ensure the data's correctness and to gain a greater understanding of the data. There was no evidence to judge whether or not: limitations of each analytical procedure, including failure to meet assumptions, were reported; multiple analytic procedures were employed to check on consistency and replicability of findings; variability, as well as central tendencies, were examined; outliers and their correctness were identified, examined and verified; statistical interactions were identified and analysed; statistical significance and practical significance was assessed; and, visual displays were used to clarify the presentation and interpretation of statistical results. With a compliance level of two requirements out of ten, the strength of the evaluation is poor.

Systematic Information (AS7): The results of the assessment indicated that protocols for quality control of the evaluation information were established, data tables generated from computer output or other means were proofread and verified; and storage of the evaluation information was systematised and controlled.

There was no evidence in the evaluation report, which suggested that evaluation staff were trained to adhere to the data protocols; that the accuracy of scoring and coding was systematically checked; whether multiple evaluators and checking the consistency of their work were used when feasible or whether data entry was verified. It was not clear from the evaluation report as to whether or not those with access to the evaluation information were defined; it was also not clear from the evaluation case whether or not access to the evaluation information according to established protocols was strictly controlled, nor was it evident from the evaluation case whether or not data providers verified the data they submitted. On this measure, the evaluation complied with three requirements out of ten. In this case the strength of the evaluation was found to be fair.

Analysis of Quantitative Information (AS8): The results showed procedures appropriate for the evaluation questions and nature of the data were chosen, and for each procedure it was specified how its key assumptions are being met.

The evaluation case, however, did not comply with eight requirements. These eight requirements included: conducting preliminary exploratory analyses to ensure the data's correctness and to gain a greater understanding of the data; reporting limitations of each analytic procedure, including failure to meet assumptions; employing multiple analytic procedures to check on consistency and replicability of findings; examining variability as well as central tendencies; identifying and examining outliers and verifying their correctness; identifying and analysing statistical interactions; assessing statistical significance and practical significance; and using visual displays to clarify the presentation and interpretation of statistical results. The strength of this evaluation was found to be poor based only two requirements being met out ten.

Analysis of Qualitative Information (AS9): Evidence showed that the evaluation focused on key questions; derived a set of categories that was sufficient to document, illuminate and respond to the evaluation questions; classified the obtained information into the validated analysis categories; and derived conclusions and recommendations, and demonstrated their meaningfulness.

However, there was no evidence, based on the contents of the evaluation case, to suggest that boundaries of information to be used were defined; information keyed to the important evaluation questions, obtained; the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders, verified; analytic procedures and methods of summarisation that were appropriate to the evaluation questions and qualitative information, chosen and employed; the derived categories for reliability and validity, tested; and limitations of the referenced information, analyses and inferences, reported. The strength of the evaluation was judged to be fair based on the compliance level of four requirements out of ten requirements.

Justified Conclusions (AS10): The results showed that the evaluation focused conclusions directly on the evaluation questions; the evaluation procedures and findings were accurately reflected; the information that supports each conclusion was cited; and the programme's side effects were identified and reported. However, there was no information found in the contents of the evaluation case to suggest that conclusions were limited to the applicable time periods, contexts, purposes and activities; neither was any information found to demonstrate that plausible alternative explanations of the findings were reported.

In addition, no explanation in the reports showed the reasons for rival explanations being rejected, nor was there any evidence of any warning against common misinterpretations or evidence that the results of a pre-release review of the draft evaluation report were obtained and addressed; nor whether or not the evaluation's limitation was reported. Here, the evaluation complied with four requirements out ten and the strength of the evaluation was found to be fair.

Impartial Reporting (AS11): The results showed that appropriate editorial authority was established; the client was engaged to determine steps to ensure fair, impartial reports; and reports were safeguarded from deliberate or inadvertent distortions. However, the contents of the evaluation case did not provide sufficient information to suggest that right-to-know audiences were determined, nor whether or not appropriate plans for releasing findings to all right-to-know audiences were established and followed.

The results did not show whether or not perspectives of all stakeholder groups were reported; whether alternative plausible conclusions were reported; whether outside audits of reports were obtained nor were steps taken to control bias described. No evidence obtained from the evaluation case suggested that there was participation in public presentations of the findings to help guard against and correct distortions by other interested parties. On this standard, the evaluation case under review complied with three requirements out of ten. Therefore, the strength of this evaluation was found to be fair.

Metaevaluation (AS12): The results under this standard showed that responsibility was assigned to someone for documenting and assessing the evaluation process and products. However, insufficient information was obtained in the evaluation case to provide substantial grounds to demonstrate that the standards to be used in judging the evaluation were designated or defined. In addition, there was no information in the evaluation case, which suggested that both formative and summative Metaevaluation were employed; or whether or not the budget was appropriate and sufficient for conducting the Metaevaluation. There was also no evidence in the evaluation case under review suggesting that a full range of information needed to judge the evaluation against the stipulated standards was recorded, nor whether an independent Metaevaluation was feasibly contracted. Furthermore, there was no evidence, which suggested that audiences receiving the Metaevaluation report were determined and recorded; nor that the instrumentation, data collection, data handling, coding and analysis against the relevant standards were evaluated.

There was a lack of clarity in the evaluation case, which demonstrated that the evaluation's involvement in, and communication of, findings to stakeholders against the relevant standards was evaluated, and whether or not a record of all Metaevaluation steps, information and analyses was maintained. With the compliance of one out of ten requirements, the strength of the evaluation case on this standard was poor. In summary the overall strength of the evaluation case study on the accuracy standards is fair.

4.1.3 The Third Evaluation Case Study.

4.1.3.1 Findings on Utility Standards (US1-US7).

The findings from the evaluation standards are described below.

Stakeholder Identification (US1). The third evaluation case complied with seven out of ten requirements under this measure. The results showed that the evaluation client was clearly identified; engaged leadership figures to identify other stakeholders; used stakeholders to identify other stakeholders; consulted potential stakeholders to identify their information needs; used stakeholders to identify other stakeholders; arranged to involve stakeholders throughout the evaluation and addressed stakeholders' evaluation needs. There was no evidence, however, in this third evaluation case under consideration, to suggest that, with the client, stakeholders were ranked for relative importance; that the evaluation was kept open to serve newly-identified stakeholders, that an appropriate range of individual stakeholders was served and that an appropriate range of stakeholder organisations was also served. Seven requirements were met under this standard and the strength of the evaluation case was found to be good.

Evaluator Credibility (US2): Under this associated standard, three out of ten requirements were met. The results showed that competent evaluators were engaged, helped stakeholders to understand the evaluation plan and stayed abreast of social and political forces.

However, there was no evidence in the evaluation report to suggest that the evaluation process engaged evaluators whom the stakeholders trust. There was also no evidence to indicate that evaluators who could address stakeholders' concerns were engaged; nor was there any evidence suggesting that evaluators who were appropriately responsive to issues of gender, socio-economic status, race and language, and cultural differences were engaged; there was no evidence of assurance that the evaluation plan responded to key stakeholders' concerns, or that stakeholders were given information on the evaluation plan's technical quality and practicality; that stakeholders' criticisms and suggestions were appropriately attended to or that interested parties were kept informed about the evaluation's progress. The strength of this evaluation case was adjudged to be fair with a compliance level of three out of ten requirements on this associated standard.

Information Scope and Selection (US3): Three requirements out of ten were met under this measure. The selection and scope of information showed that the evaluation assigned priority to the most important stakeholders; obtained sufficient information to assess the programme's merit and also obtained sufficient information to assess the programme's worth.

However, the evaluation did not comply with the remaining seven requirements. There was no evidence in the evaluation case, which suggested that the client's most important evaluation requirements were understood; that stakeholders were interviewed to determine their different perspectives; that there was assurance that the evaluator and client negotiated pertinent audiences, questions and required information; that priority was assigned to the most important questions; that there was allowance for flexibility in adding questions during the evaluation; that sufficient information was obtained to address the stakeholders' most important evaluation questions; and that the evaluation effort was allocated in accordance with the priorities assigned to the necessary information. Therefore, with compliance of three out ten requirements, the strength of the evaluation case on this associated standard was found to be fair.

Values Identification (US4): The third evaluation case complied with seven out of ten requirements under this standard. The evidence reviewed showed that the evaluation under consideration considered alternative sources of values for interpreting evaluation findings; provided a clear, defensible basis for value judgments; identified pertinent societal needs; identified pertinent customer needs; referenced, as appropriate, the relevant institutional mission; referenced the programme's goals; and, as appropriate, presented alternative interpretations based on conflicting but credible value bases.

There were three requirements that were not met under this secondary standard. The outcomes of the assessment showed that appropriate party(ies) and alternative sources of values for interpreting evaluation findings were not considered; the stakeholders' values were not taken into account; and pertinent laws were not referenced. The strength of this evaluation case was judged to be very good with a compliance level of seven out of ten requirements on this associated standard.

Report Clarity (US5): On the basis of this measure, all ten requirements were met. An examination of the contents of the evaluation case, guided by the prescriptions in the PES, confirmed the following: essential information was clearly reported; the reports were brief, simple and direct; the reports focused on contracted questions; the programme and its contents were described; the purposes, procedures and findings of the evaluation were also described; the conclusions and recommendations were supported; reporting in technical jargon was avoided; an executive summary was provided; the evaluation was reported in the language(s) of stakeholders; and a technical report was provided. With a compliance level of ten requirements out ten, the strength of the evaluation cases was found to be excellent on this associated standard.

Report Timeliness and Dissemination (US6): The third evaluation case complied with five requirements out of ten under this measure. The assessment of the reports showed that timely interim reports to intended users were made; the final report was delivered when needed; there were timely exchanges with the programme's policy board; the presentations were kept appropriately brief; and examples to help audiences relate the findings to practical situations were used. However, based on the contents of the evaluation, there was no evidence in the evaluation case, which suggested that timely exchanges were held with the programme's staff; nor were they held with the programme's customers; nor with the public media; nor with the full range of right-to-know audiences; nor were that effective media employed for reaching and informing the different audiences. The strength of the evaluation case was good as it complied with five out of ten requirements under this associated standard.

Evaluation Impact (US7): On the basis of this measure, it was found from the assessment of the report that seven out of ten requirements were met. The findings revealed that stakeholders were involved throughout the evaluation; that the evaluation process encouraged and supported stakeholders' use of the findings; and that stakeholders were shown how they might use the findings in their work. Results of the scrutiny demonstrated that potential uses of findings were forecast and addressed; that interim reports were provided; that reports were open, frank and concrete; and that arrangements were made to provide follow-up assistance in interpreting and applying the findings. On the other hand, there was no evidence in the evaluation case suggesting that contact with audience was maintained; that written reports were supplemented with ongoing oral communication; nor that feedback workshops to go over and apply findings were conducted.

With compliance levels of seven requirements out of ten, the strength of the evaluation was found to be very good. In summary, with four very good scores, one fair score, one good score and an excellent score under the utility standard, overall the strength of the evaluation case on this standard was fair.

4.1.3.2. Findings on Feasibility Standards (FS1-FS3).

The specific patterns of the findings that emerged for each of the above associated evaluation standard are described below.

Practical Procedures (FS1): The first evaluation case complied with eight out of ten requirements under this measure. Evidence generated from the assessment showed that the evaluation under consideration tailored the methods and instruments to information requirements; the data burden and disruption were both minimised; competent staff were appointed; staff were trained; procedures were chosen that the staff were qualified to carry out; procedures, in light of known constraints, were chosen; and a realistic schedule was made. However, there was no evidence from the examination of the evaluation case showing that local interested parties including beneficiaries were engaged to help to conduct the evaluation. It is not clear from the evaluation that evaluation procedures were appropriately made part of routine events. The evaluation complied with eight requirements out of ten; therefore, its strength was judged to be very good.

Political Viability (FS2): The third evaluation case complied with eight out of ten requirements under this measure. Results generated from the assessment showed that cooperation was fostered; there was agreement on the editorial and dissemination authority; interim reports were issued; divergent views were reported; and a firm public contract was employed.

However, it is not clear whether differing positions of different interest groups were anticipated; there was no information found showing: that attempts to bias or misapplying the findings were averted or counteracted; whether or not, stakeholders throughout the evaluation where involved, whether or not, right-to-know audiences were reported to; or whether or not, any corrupted evaluation was terminated. The evaluation case complied with five requirements out of ten. Therefore, the strength of the evaluation case on this measure was found to be good.

Cost-effectiveness (FS3): The third evaluation case complied with six requirements out of ten requirements under this measure. The assessment of the evaluation report showed that the evaluation was efficient; informed decisions; fostered programme improvement; provided accountability information; generated new insights and helped to spread effective practices. However, there was no evidence in the evaluation case suggesting that the evaluation process produced information worth the investment or minimised disruptions and time demands on programme personnel. The evaluation case under consideration complied with seven requirements out of ten and the strength of the case on this measure was found to be very good. With two scores of good under the feasibility standard, the overall strength of the evaluation case on this standard is found to be good. In summary, the overall strength of the third evaluation case on the feasibility standard is found to be good.

4.1.3.3. Propriety Standards (PS1-PS8).

The specific patterns of the findings that emerged for each of the above associated evaluation standard are described below.

Service Orientation (PS1): The results showed that the evaluation assessed the needs of the programme's customers; assessed programme outcomes against targeted customers' assessed needs; helped to ensure that the full range of rightful programme beneficiaries are served; promoted excellent service; clarified the evaluation's service orientation for stakeholders; identified programme strengths on which to build; identified programme weaknesses that required correction; and informed all right-to-know audiences of the programme's positive and negative outcomes. By contrast, there is no evidence to suggest that the evaluation gave interim feedback for programme improvement or exposing harmful practices. On this measure alone, the evaluation complied with eight requirements out of ten. The strength of the evaluation was, therefore, judged to be very good.

Formal Agreements (PS2): The results showed that formal parties agreed on evaluation reports, editing, release of reports, evaluation staff and evaluation resources. The results of the assessment, however, reveal that formal parties did not agree on the confidentiality/anonymity of data, evaluation purpose and questions, evaluation procedures and schedule nor Metaevaluation. The strength of the evaluation was found to be good because it complied with five out ten requirements.

Rights of Human Subjects (PS3): The evidence generated demonstrated that the process of designing and conducting the evaluation clarified its intended uses; followed due process; upheld civil rights; understood participant values, followed protocol; honoured confidentiality/anonymity agreements and did no harm. The examination of the evaluation case showed that the designing and conducting of the evaluation did not, however, make clear to stakeholders that the evaluation would respect and protect the rights of human subjects; stakeholders were not kept informed and there was no evidence to suggest that they respected diversity. On this measure, the evaluation complied with seven out of ten requirements. The overall strength of the evaluation was found to be very good.

Human Interactions (PS4): On this measure, the results showed that evaluators consistently related to all stakeholders in a professional manner; maintained effective communication with stakeholders; followed the institution's protocol; minimised disruption; and honoured participants' privacy rights and time commitments. There is also no evidence in the evaluation case, which suggested that the evaluation ignored or helped to cover up any participant's incompetence, unethical behaviour, fraud, waste or abuse.

There was no evidence in the evaluation case suggesting that evaluators were alert to or addressed participants' concerns regarding the evaluation; there was also no evidence to demonstrate whether or not the evaluation process was sensitive to participants' diversity of values and cultural differences; or whether or not the evaluation process was even-handed in addressing different stakeholders. The evaluation on this measure was judged to be good based on its compliance with six out of ten requirements.

Complete and Fair Assessment (PS5): The evidence generated demonstrated that the evaluation case complied with six out of ten requirements. The results showed that the evaluation case assessed and reported the programme's strengths; assessed and reported its weaknesses; reported on intended outcomes; reported on unintended outcomes; had the draft report reviewed; and acknowledged the final report's limitations.

Nevertheless, the evaluation case did not give a thorough account of the evaluation's process, as appropriate; nor did it show how the programme's strengths could be used to overcome its weaknesses. The effects of the evaluation's limitations on the overall judgment of the programme were not estimated or reported. In addition, there was no evidence suggesting that criticisms of the draft report were appropriately addressed. On this associated standard, the strength of the evaluation was found to be good because it complied with six out ten requirements.

Disclosure of Findings (PS6): On this measure the evaluation complied with four requirements out of ten and the strength of the evaluation was judged to be fair. In this case, results indicated that formal parties defined the right-to-know audiences; reported all findings in writing; showed the basis for the conclusions and recommendations; and in reporting, adhered strictly to a code of directness, openness and completeness. However, the evaluation did not establish a contractual basis for complying with right-to-know requirements; did not report relevant points of view of both supporters and critics of the programme; did not disclose the evaluation's limitations; and did not ensure that reports reach their audiences.

Conflict of interest (PS7): In the third evaluation case, the results showed that in dealing with conflicts of interest, the evaluation process engaged multiple evaluators; maintained evaluation records for independent review, when appropriate, released evaluation procedures, data and reports for public review; and contracted with the funding authority rather than the funded programme.

However, the evidence based on the contents of the evaluation case showed that in dealing with conflicts of interest, the evaluation process did not identify potential conflicts of interest early in the evaluation; did not provide written, contractual safeguards against identified conflicts of interest; did not, as appropriate, engage independent parties to assess the evaluation for its susceptibility or corruption by conflicts of interest; did not have internal evaluators report directly to the chief executive Officer; did not report equitably to all right-to-know audiences, did not engage uniquely-qualified persons to participate in the evaluation, even if they had a potential conflict of interest; and could not take steps to counteract the conflict. On this measure, the evaluation complied with four requirements out of ten and the strength of the evaluation was judged to be fair.

Fiscal Responsibility (PS8): The results under this standard indicated that the evaluation specified and budgeted for expense items in advance; kept the budget sufficiently flexible to permit appropriate re-allocations to strengthen the evaluation; obtained appropriate approval for necessary budgetary modifications; assigned responsibility for managing the evaluation finances; moreover, evidence from the evaluation case under consideration suggested that the evaluation process was frugal in expending evaluation resources.

There was no evidence provided in the evaluation case, which indicated that accurate records of sources of funding and expenditures were maintained; that comparison shopping for evaluation materials was employed; that comparison contract bidding was employed; or that, as appropriate, an expenditure summary as part of the public evaluation report was included. On this measure, the evaluation complied with five requirements out of ten and the strength of the evaluation was judged to be good. In summary, the overall strength of the evaluation case on this standard was found to be good.

4.1.3.4. Findings on Accuracy Standards (AS1-AS12).

The patterns of the findings that emerged under these evaluation standards are described below.

Programme Documentation (AS1): The evidence generated from assessing the evaluation case demonstrated that descriptions of the intended programme from various written sources were collected; the evaluation: described the way in which the programme was intended to function; showed evidence of maintained records from various sources of the way in which the programme operated; described how the programme actually functioned; analysed discrepancies between the various descriptions of how the programme was intended to function; analysed discrepancies between the way in which the programme was intended to operate and how it actually operated; and produced a technical report that documented the programme's operations.

However, there was no evidence in the evaluation to indicate that descriptions of the intended programme were collected from the client and various stakeholders; that the process engaged independent observers, as feasible, to describe the programme's actual operations; that the client and various stakeholders were asked to assess the accuracy of recorded descriptions of both the intended and the actual programme. On this measure, the evaluation complied with seven requirements out of ten and the strength of the evaluation was judged to be very good.

Context Analysis (AS2): An examination of the evaluation case indicated that multiple sources of information were used to describe the programme's context; that the context's technical, social, political, organisational and economic features were described; that a log of unusual circumstances was maintained; the way in which the programme's context was similar to, or different from, contexts where the programme might be adopted was analysed; and the effects of context on programme outcomes were estimated.

However, there was no evidence in the report, which indicated that instances in which individuals or groups, intentionally or otherwise, interfered with the programme were recorded; instances in which individuals or groups, intentionally or otherwise, gave special assistance to the programme were recorded; that those contextual influences that appeared to influence the programme significantly and which might have been of interest to potential adopters were reported; that any critical competitors to this programme that functioned at the same time and in the programme's environment were identified and described; or that the way in which people in the programme's general area perceived the programme's existence, importance and quality was described. With compliance on five requirements out of ten, the strength of the evaluation case on this standard was found to be good.

Describe Purposes and Procedures (AS3): The evidence from the assessment showed that at the evaluation's outset the client's purposes for the evaluation were recorded; stakeholders' intended uses of evaluation findings were monitored and described; and the actual evaluation procedures, as implemented, were recorded. However, there was no evidence to enable the researcher to make plausible judgment as to whether or not the evaluation's purposes stayed the same or changed over time; whether or not points of agreement and disagreement among stakeholders regarding the valuation's purposes were identified and assessed; or whether or not evaluation procedures to accommodate changes in the evaluation's purposes were, as appropriate, updated.

Furthermore, there was also no evidence in the evaluation case that demonstrated whether or not when interpreting findings, the evaluation process took into account the extent to which the intended procedures were effectively executed; whether or not the evaluation's purposes and procedures in the summary and full-length evaluation reports were described; or whether or not independent evaluators monitored and evaluated the evaluation's purposes and procedures. On this measure, the evaluation met three out ten requirements and the strength of the evaluation was found to be fair.

Defensible Information Sources (AS4): On this measure, the results showed that the evaluation complied with four out of ten requirements. The assessment of the evaluation showed that a variety of data collection methods were, as appropriate, employed; that the criteria and methods used to select information sources were documented, justified and reported; that data collection instruments were included in a technical appendix to the evaluation report; and that any biasing features in the obtained information were documented and reported.

However, the results of the assessment of the evaluation could not provide sufficient information to determine whether or not the information from a variety of sources was obtained; whether or not pertinent, previously-collected information, once validated, was used; whether or not information sources were documented and reported; whether or not, for each source, the population was defined; whether or not, for each population, any employed sample was, as appropriate, defined; or whether or not the means used to obtain information from each source was documented, justified and reported. The evaluation report complied with four requirements for this standard, therefore, the strength of the evaluation was judged to be fair.

Valid Information (AS5): On the basis of this measure, evidence suggested that the evaluation process focused the evaluation on key questions; provided a detailed description of the constructs and behaviours regarding the information that was acquired; the data collectors were trained and calibrated; the way in which information from each procedure was scored, analyzed, and interpreted was documented; that the evaluation assessed and reported the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions; and that the evaluation established meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures.

However, there was no evidence in the evaluation case under consideration which suggested that multiple measures to address each question were, as appropriate, employed; whether the type of information each employed procedure acquired was assessed and reported; whether or not the data collection conditions and process were documented and reported; and, finally, whether or not inferences, singly or in combination, were reported and justified. On this standard, six requirements out of ten were met. The strength of the evaluation was, therefore, good.

Reliable Information (AS6): The evaluation report complied with two requirements under this associated criteria. The results showed that for each employed data collection device, the unit of analysis was specified and that reliability problems in the final report were acknowledged. However, the evaluation case did not comply with the following eight requirements: there was no evidence in the evaluation case to ascertain that preliminary exploratory analyses were conducted to ensure the data's correctness and to gain a greater understanding of the data.

There was no evidence to adjudge that limitations of each analytic procedure, including failure to meet assumptions, were reported; that multiple analytic procedures were employed to check on consistency and replicability of findings; that variability, as well as central tendencies, were examined, that outliers and their correctness were identified, examined and verified; that statistical interactions were identified and analysed. That statistical and practical significance was assessed, or that visual displays were used to clarify the presentation and interpretation of statistical results. With the compliance level of two requirements out of ten, the strength of the evaluation was poor.

Systematic Information (AS7): The results of the assessment indicated that protocols for quality control of the evaluation information were established; data tables generated from computer output or other means were proofread and verified; and storage of the evaluation information was systematised and controlled.

There was no evidence in the evaluation report under consideration to suggest that evaluation staff were trained to adhere to the data protocols; the accuracy of scoring and coding were systematically checked; when feasible, used multiple evaluators and checked the consistency of their work; verified data entry; defined who would have access to the evaluation information; strictly controlled access to the evaluation information according to established protocols or had data providers verify the data they submitted**.** On this measure, the evaluation complied with three requirements out of ten. In this case the strength of the evaluation was found to be fair.

Analysis of Quantitative Information (AS8): The results showed that procedures appropriate for the evaluation questions and nature of the data were chosen, and for each procedure it was specified how its key assumptions were met. The evaluation case, however, did not comply with eight requirements. These eight requirements included: conducting preliminary exploratory analyses to ensure the data's correctness and to gain a greater understanding of the data; reporting limitations of each analytic procedure, including failure to meet assumptions; employing multiple analytic procedures to check on consistency and replicability of findings; examining variability, as well as central tendencies; identifying and examining outliers and verifying their correctness; identifying and analysing statistical interactions; assessing statistical significance and practical significance; and using visual displays to clarify the presentation and interpretation of statistical results. The strength of this evaluation was found to be poor based only two out ten requirements being met.

Analysis of qualitative information (AS9): Evidence showed that the evaluation focused on key questions; derived a set of categories that were sufficient to document, illuminate and respond to the evaluation questions; classified the obtained information into the validated analysis categories; and derived conclusions and recommendations, and demonstrated their meaningfulness.

There was also no evidence based on the contents of the evaluation case that: suggested boundaries of information to be used were defined; information key to the important evaluation questions was obtained; the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders, was verified; analytic procedures and methods of summarising that were appropriate to the evaluation questions were employed; qualitative information was chosen; the process derived and tested categories for reliability and validity; and limitations of the referenced information, analyses and inferences were reported. The strength of the evaluation was adjudged to be fair based on the compliance level of four out of ten requirements.

Justified Conclusions (AS10): The results showed that: the evaluation focused conclusions directly on the evaluation questions; the evaluation procedures and findings were accurately reflected; the information that supported each conclusion was cited; and the programme's side effects were identified and reported.

However, there was no information found in the contents of the evaluation case that suggested that conclusions were limited to the applicable time periods, contexts, purposes and activities; nor that plausible alternative explanations of the findings were reported. There was no explanation in the reports to show why rival explanations were rejected; no evidence of any warning against common misinterpretations; no evidence that the results of a pre-release review of the draft evaluation report were obtained and addressed or whether or not evaluations limitation was reported. Here the evaluation complied with four requirements out ten and the strength of the evaluation was found to be fair.

Impartial Reporting (AS11): The results showed that appropriate editorial authority was established; that the client was engaged to determine steps to ensure fair, impartial reports; and that reports were safeguarded from deliberate or inadvertent distortions. However, the contents of the evaluation case did not provide sufficient information to suggest that: right-to-know audiences were determined; appropriate plans for releasing findings to all right-to-know audiences were established and followed; perspectives of all stakeholder groups were reported; alternative plausible conclusions were reported; outside audits of reports were obtained; steps were taken to control bias; and that there was participation in public presentations of the findings to help to guard against and correct distortions by other interested parties. On this standard, the evaluation complied with three requirements out of ten. Therefore, the strength of this evaluation was found to be fair.

Metaevaluation (AS12): The results under this standard showed that responsibility was assigned to someone for documenting and assessing the evaluation process and products. However, sufficient information was not obtained in the evaluation case, to provide substantial grounds to demonstrate that the standards used in judging the evaluation were designated or defined; whether or not both formative and summative Metaevaluation was employed; or whether or not the budget was appropriate and sufficient for conducting the Metaevaluation.

In addition, there was no evidence suggesting that a full range of information necessary to judge the evaluation against the stipulated standards was recorded; that an independent Metaevaluation was feasibly contracted; or whether or not audiences that received the Metaevaluation report were determined and recorded. There was no indication as to whether or not the instrumentation, data collection, data handling, coding and analysis against the relevant standards were evaluated; whether or not the evaluation's involvement in, and communication of, findings to stakeholders against the relevant standards were evaluated; or whether or not a record of all Metaevaluation steps, information and analyses were maintained. With the compliance of only one out ten requirements, the strength of the evaluation of this standard was poor. In summary, the overall strength of the evaluation case study on the accuracy standard was fair.

4.1.4 The Fourth Evaluation Case.

4.1.4.1 Findings on Utility Standards (US1-US7).

The findings that emerged for each of the above associated evaluation standard are described below.

Stakeholder Identification (US1): The fourth evaluation case complied with two out of ten requirements under this measure. The results showed that the evaluation client was clearly identified and that potential stakeholders were consulted in order to identify their information needs. However, the results under this measure demonstrated that the evaluation case did not comply with eight requirements. There was no evidence in this evaluation case to show that leadership figures were engaged to identify other stakeholders.

In addition, stakeholders were not used to identify other stakeholders; stakeholders were not ranked for relative importance with the client; there was no arrangement made to involve stakeholders throughout the evaluation; the evaluation was not kept open to serve newly-identified stakeholders; and the evaluation did not address the stakeholders' needs. There was no evidence in the evaluation case to enable the researcher to judge whether or not both appropriate ranges of individual stakeholders and organisations were served. On this measure alone, the strength of the evaluation case was found to be poor.

Evaluator Credibility (US2): Two out of ten requirements were met under this associated standard. The results showed that competent evaluators were engaged and that evaluators gave stakeholders information on the evaluation plan's technical quality and practicality. However, the evaluation did not comply with six requirements. It was not clear in the evaluation case under consideration that evaluators, whom the stakeholders trusted, were engaged. Nor was it clear if the evaluators engaged were appropriately responsive to gender issues, socio-economic status, race and language, or cultural differences.

There was also no evidence in the evaluation case to suggest that the evaluation plan responded to key stakeholders' concerns or whether or not stakeholders were helped to understand the evaluation plan. In addition, results of the assessment did not show that evaluators attended appropriately to stakeholders' criticisms and suggestions. Finally, no evidence in the report suggested that evaluators stayed abreast of social and political forces or that evaluators kept some interested parties informed about the evaluation's progress. The strength of this evaluation case was judged to be poor with a compliance level of only two out of ten requirements with this associated standard.

Information Scope and Selection (US3): Five out of ten requirements were met under this measure. The selection and scope of information showed that the client's most important evaluation requirements were understood; that priority was assigned to the most important stakeholders; that priority was assigned to the most important questions; that information obtained was sufficient to address the stakeholders' most important evaluation question; and that evaluation effort was allocated in accordance with the priorities assigned to the necessary information.

However, the evaluation did not comply with the other five requirements. Results of the assessment showed that stakeholders were not interviewed to determine their different perspectives, that there was no assurance that the evaluator and client negotiated with pertinent audiences, questions and required information; that flexibility was allowed for adding questions during the evaluation; that sufficient information was obtained to assess the programme's merit; or that sufficient information was obtained to assess the programme's worth. Therefore, with compliance of five out ten requirements, the strength of the evaluation case on this associated standard was found to be good.

Values identification (US4): The fourth evaluation case complied with four out of ten requirements under this measure. Evidence showed that the evaluation under consideration provided a clear, defensible basis for value judgments; that the evaluation determined the appropriate party(ies) to make the evaluational interpretations; that reference was appropriately made regarding the relevant institutional mission; and that reference was made regarding the programme's goals.

However, six requirements were not met under this standard. The outcomes of the assessment showed that: alternative sources of values for interpreting evaluation findings were not considered; pertinent societal needs were not identified; pertinent customer needs were not identified; reference was not made to pertinent laws; stakeholders' values were not taken into account; and alternative interpretations were not appropriately presented based on conflicting but credible values. The strength of this evaluation case was judged to be fair with the compliance level of four out of ten requirements with this associated standard.

Report Clarity (US5): On the basis of this measure, all ten requirements were met. Examination of the contents of the evaluation case, guided by the prescriptions in the PES, confirmed the following: essential information was clearly reported; the reports were brief, simple and direct; the reports focused on contracted questions; the programme and its contents were described; the purposes, procedures and findings of the evaluation were described; conclusions and recommendations were supported; reporting in technical jargon was avoided; an executive summary was provided; the evaluation was reported in the stakeholders' language(s); and a technical report was provided. With a compliance level of ten requirements out ten, the strength of the evaluation case was found to be excellent with this associated standard.

Report Timeliness and Dissemination (US6): The fourth evaluation case complied with seven out of ten requirements under this measure. The assessment of the reports showed that the interim reports were made on a timely basis to intended users; the final report was delivered when it was needed; there were timely exchanges with the Programme's policy board; there were timely exchanges with its staff; there were timely exchanges with the full range of right-to-know audiences; effective media were employed for reaching and informing the different audiences; and the presentations was kept appropriately brief. However, the assessment of the evaluation case showed that timely exchanges were not made with both the programme's customers and the public media, and the audience was not helped to relate the findings to practical situations. The strength of the evaluation case was very good as it complied with seven requirements under this standard.

Evaluation Impact (US7): On the basis of this measure, it was found from the assessment of the report that four out of ten requirements were met. Evidence generated from the review revealed that contact was maintained with the audience; stakeholders' use of the findings was encouraged and supported; interim reports were provided; and reports were open, frank and concrete.

On the other hand, stakeholders were not involved throughout the evaluation, nor were they shown how they might use the findings in their work; potential uses of the findings were not forecast and addressed; written reports were not supplemented with ongoing oral communication; there was no evidence in the evaluation case suggesting that feedback workshops were conducted on how to apply findings; and arrangements were not made to provide follow-up assistance in interpreting and applying the findings.

With compliance levels of four requirements out of ten, the strength of the evaluation was found to be fair. With three fair scores, one each poor, good, very good and excellent under the utility standard, overall the strength of the evaluation case on this standard is predominantly fair.

4.1.4.2 Findings on Feasibility Standards (FS1-FS3).

The specific findings that emerged for each of the above associated evaluation standard are presented below.

Practical Procedures (FS1): The second evaluation case complied with eight out of ten requirements under this measure. Evidence generated from the assessment showed that the evaluation under consideration tailored the methods and instruments to information requirements; that the data burden was minimised; that competent staff were appointed; staff were trained; procedures were chosen that the staff are qualified to carry out; procedures in light of known constraints were chosen; and a realistic schedule was made. However, there was no evidence showed that disruption was minimised nor were locals engaged to help conduct the evaluation. It was not clear from the evaluation whether evaluation procedures were appropriately made part of routine events. The evaluation complied with seven requirements out of ten. Therefore, its strength was judged to be very good.

Political Viability (FS2). The fourth evaluation case complied with eight out of ten requirements under this measure. Evidence generated from the assessment showed that cooperation was fostered; there was agreement on the editorial and dissemination authority; interim reports were issued; and a firm public contract was employed.

However, the results of the assessment also revealed that stakeholders were not involved throughout the evaluation and divergent views were not reported. In addition, it was not clear whether all right-to-know audiences received the report; it is also not clear whether any corrupted evaluations were terminated. Moreover, the evaluation case does not indicate whether differing positions of different interest groups were anticipated; and, finally, no evidence in the evaluation case suggested that attempts of bias or misapplication of the findings were averted or counteracted. The evaluation case complied with four requirements out of ten. Therefore, the strength of the evaluation case on this measure was found to be fair.

Cost-effectiveness (FS3). The third evaluation case complied with six out of ten requirements under this measure. The assessment of the evaluation report showed that the evaluation was: efficient; informed decisions; fostered programme improvement; provided accountability information; generated new insights; and helped to spread effective practices.

However, there was no evidence in the evaluation case to suggest that the evaluation process made use of in-kind services; produced information worth the investment; or minimised disruptions and time demands on programme personnel. The evaluation case complied with seven requirements out of ten and the strength of the case on this measure was found to be very good. With two scores of good under the feasibility standard, the overall strength of the evaluation case on this standard was found to be good.

4.1.4.3 Propriety Standards (PS1-PS8): The findings that emerged for each of the above associated evaluation standard are described below.

Service Orientation (PS1): The results showed that the evaluation assessed the needs of the programme's customers; assessed programme outcomes against targeted customers' assessed needs; helped to ensure that the full range of rightful programme beneficiaries are served; promoted excellent service; made the evaluation's service orientation clear to stakeholders; identified programme strengths on which to build; identified programme weaknesses to correct; and informed all right-to-know audiences of the programme's positive and negative outcomes. By contrast, there was no evidence suggesting that the evaluation gave interim feedback for programme improvement or exposed harmful practices. On this measure alone, the evaluation complied with eight requirements out of ten. The strength of the evaluation was, therefore, judged to be very good.

Formal Agreements (PS2). The results showed that formal parties agreed on the evaluation purpose and questions, audiences, evaluation reports, editing, release of reports, evaluation procedures and schedule, evaluation staff and evaluation resources. The results of the assessment, however, revealed that formal parties did not agree on both the confidentiality/anonymity of data or Metaevaluation. The strength of the evaluation was found to be very good because it complied with eight out of ten requirements.

Rights of Human Subjects (PS3): The evidence generated demonstrated that the process of designing and conducting the evaluation made clear to stakeholders that the evaluation would respect and protect the rights of human subjects; clarified intended uses of the evaluation; followed due process; understood participant values; followed protocol; and honoured confidentiality/anonymity agreements. The examination of the evaluation case showed that the designing and conducting of the evaluation did not, however, keep stakeholders informed; did not uphold civil rights; did not respect diversity; and, based on the contents of the evaluation, it is not clear whether or not the valuation did not do any harm. On this measure, the evaluation complied with six requirements and the overall strength of the evaluation was found to be good.

Human Interactions (PS4): On this measure, the results showed that evaluators consistently related to all stakeholders in a professional manner; maintained effective communication with stakeholders; followed the institution's protocol; minimised disruption; honoured participants' privacy rights; and honoured time commitments. There was also no evidence in this evaluation report, which suggested that the evaluation ignored or helped to cover up any participant's incompetence, unethical behaviour, fraud, waste or abuse. However, no evidence in the evaluation case suggested that evaluators were alert to or addressed participants' concerns about the evaluation; whether or not they were sensitive to participants' diversity of values and cultural differences; and whether or not they were even-handed in addressing different stakeholders. The evaluation on this measure was judged to be good, based on its compliance with six out of ten requirements.

Complete and Fair Assessment (PS5): The evidence generated demonstrated that the evaluation case complied with five out of ten requirements .The results showed that the evaluation case assessed and reported the programme's strengths; assessed and reported its weaknesses; reported on intended outcomes; gave a thorough account of the evaluation's process; and had the draft report reviewed.

Nevertheless, the evaluation did not report on unintended outcomes; did not, as appropriate, show how the programme's strengths could be used to overcome its weaknesses; did not appropriately address criticisms of the draft report; did not acknowledge the final report's limitations; and did not estimate and report the effects of the evaluation's limitations on the overall judgment of the programme. The strength of the evaluation was found to be good because it complied with five out of ten requirements.

Disclosure of Findings (PS6): On this measure the evaluation complied with four requirements out of ten and the strength of the evaluation was judged to be fair. In this case, results indicated that the formal parties defined the right-to-know audiences; reported all findings in writing; showed the basis for the conclusions and recommendations; and in reporting, adhered strictly to a code of directness, openness and completeness. However, the evaluation did not establish a contractual basis for complying with right-to-know requirements; did not report relevant points of view of both supporters and critics of the programme; did not disclose the evaluation's limitations; in reporting, did not adhere strictly to a code of directness, openness and completeness; and did not ensure that reports reached their audiences.

Conflict of Interest (PS7): In the case of the second evaluation case, the results showed that in dealing with conflict of interest, the evaluation process engaged multiple evaluators; maintained evaluation records for independent review; when appropriate, released evaluation procedures, data and reports for public review; and contracted with the funding authority rather than the funded programme.

However, the evidence based on the contents of the evaluation case showed that in dealing with conflicts of interest, the evaluation process did not identify potential such conflicts early in the evaluation; did not provide written, contractual safeguards against identified conflicts of interest; did not, as appropriate, engage independent parties to assess the evaluation for its susceptibility or corruption by conflicts of interest; had internal evaluators report directly to the chief executive officer; did not report equitably to all right-to-know audiences; did not engage uniquely-qualified persons to participate in the evaluation, even if they had a potential conflict of interest; and could not take steps to counteract the conflict. On this measure, the evaluation complied with four requirements out of ten and the strength of the evaluation was judged to be fair.

Fiscal Responsibility (PS8): The results under this standard indicated that the evaluation specified and budgeted for expense items in advance; kept the budget sufficiently flexible to permit appropriate re-allocations to strengthen the evaluation; obtained appropriate approval for necessary budgetary modifications; assigned responsibility for managing the evaluation finances; and evidence from the assessment suggested that the evaluation process was frugal in expending evaluation resources.

There was no evidence provided in the evaluation case, which indicated that accurate records of sources of funding and expenditure were maintained; that comparison shopping for evaluation materials was employed; that comparison contract bidding was employed; or that, as appropriate, an expenditure summary as part of the public evaluation report was included. On this measure, the evaluation complied with five requirements out of ten and the strength of the evaluation was judged to be good. In summary, the overall strength of the evaluation case on this standard was found to be good.

4.1.4.4. Accuracy Standards (AS1-AS12).

Patterns of findings that emerged for each of the above associated evaluation standard are described below.

Programme Documentation (AS1): The evidence generated from assessing the evaluation demonstrated that descriptions of the intended programme from various written sources were collected; the way the programme was intended to function was described; records from various sources of the way the programme operated were maintained; the way the programme actually functioned was described; discrepancies between the various descriptions of the way the programme was intended to function were analysed; discrepancies between the way the programme was intended to operate and the way it actually operated were analysed, and a technical report that documents the programme's operations was produced.

However, there was no evidence in the evaluation, which indicated whether or not descriptions of the intended programme from the client and various stakeholders were collected, as feasible; that independent observers to describe the programme's actual operations were engaged; and whether or not both the client and various stakeholders were asked to assess the accuracy of recorded descriptions of both the intended, and the actual, programme. On this measure, the evaluation complied with seven requirements out of ten and the strength of the evaluation was judged to be very good.

Context Analysis (AS2): An examination of the evaluation case indicated that multiple sources of information were used to describe the programme's context; the context's technical, social, political, organisational and economic features were described; the evaluation analysed the way the programme's context is similar to, or different from, contexts where the programme might have been adopted.

However, there was no evidence in the evaluation case, which indicated that a log of unusual circumstances was maintained; nor that the evaluation recorded instances in which individuals or groups, intentionally or otherwise, interfered with the programme; neither was there any evidence that the evaluation recorded instances in which individuals or groups, intentionally or otherwise, gave special assistance to the program; and lastly there was no evidence in the evaluation case, which reported those contextual influences that appeared to influence the programme significantly and that might be of interest to potential adopters.

Furthermore, there was no evidence in the evaluation case, which indicated that the effects of context on programme outcomes were estimated or that any critical competitors to this programme that functioned at the same time were identified and described in the programme's environment; there was also no evidence in the evaluation case, which described the way people in the programme's general area perceived the programme's existence, importance and quality. With compliance of three requirements out of ten, the strength of the evaluation case on this standard was found to be fair.

Describe Purposes and Procedures (AS3): The evidence from the assessment showed that, at the evaluation's outset, the client's purposes for the evaluation were recorded; actual evaluation procedures, as implemented, were recorded; and that the evaluation's purposes and procedures in the summary and full length evaluation reports were described.

However, there was no evidence to enable the researcher to make plausible judgment as to whether stakeholders' intended uses of evaluation findings were described and monitored; no evidence which suggests how the evaluation's purposes stay the same or change over time were described or monitored; whether or not points of agreement and disagreement among stakeholders regarding the valuation's purposes were identified and assessed; whether or not, as appropriate, evaluation procedures to accommodate changes in the evaluation's purposes were updated; there was no indication in the evaluation cases that indicated that when interpreting findings, different stakeholders' intended uses of the evaluation were taken into account; also when interpreting findings, whether the extent to which the intended procedures were effectively executed were also taken into account was not indicated.

The researcher could not ascertain, based on the contents of the evaluation case under review, that independent evaluators to monitor and evaluate the evaluation's purposes and procedures were feasibly engaged. On this measure, the evaluation met three out ten requirements and the strength of the evaluation was found to be fair.

Defensible Information Sources (AS4): On this measure, the results showed that the evaluation complied with two requirements out of ten. The evaluation showed that a variety of data collection methods was appropriately used and included data collection instruments in a technical appendix to the evaluation report.

However, the results of the assessment of the evaluation could not provide sufficient information to determine whether or not the information was obtained from a variety of sources; whether or not previously collected information, once validated, was pertinent; whether or not information sources were documented and reported or that the criteria and methods used to select information sources were documented, justified and reported. The information from the review could not indicate whether or not: the population was defined for each source; any employed sample for each population was defined appropriately; the means used to obtain information from each source were documented, justified and reported; or any biasing features in the obtained information were documented and reported. The evaluation report complied with two requirements for this standard, therefore, the strength of the evaluation was judged to be poor.

Valid Information (AS5): On the basis of this measure, evidence suggested that the evaluation focused on key questions; as appropriate, employed multiple measures to address each question; and trained and calibrated data collectors.

However, there was insufficient information to judge whether or not a detailed description of the constructs and behaviours was provided, about which information was acquired; nor was there any evidence upon which to judge whether the type of information each employed procedure acquired was assessed and reported. It was not clear whether or not the collection conditions and process were documented and reported.

The evidence generated from the review could not demonstrate whether or not the way in which information from each procedure was scored, analysed, interpreted and documented; whether or not inferences, singly and in combination, were reported and justified; whether or not the comprehensiveness of the information, provided by the procedures, as a set in relation to the information needed to answer the set of evaluation questions, was assessed and reported; whether or not meaningful categories of information, by identifying regular and recurrent themes in information collected using qualitative assessment procedures, were established. On this standard, three requirements out of ten were met. The strength of the evaluation was, therefore, fair.

Reliable Information (AS6): The evaluation report complied with five requirements under these associated criteria. Evidence showed that the unit of analysis for each employed data collection device was specified; measuring devices that, in the past, have shown acceptable levels of reliability for their intended uses, were as feasibly, chosen; scorers and analysts to produce consistent results were trained and calibrated; and the consistency between multiple observers was appropriately checked and engaged.

However, the evaluation case did not comply with six requirements. Evidence showed that the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions and the evaluator’s bias in reporting the reliability of an instrument, were not assessed or reported; the consistency of scoring, categorisation and coding were not checked or reported; in order to identify and control sources of error, pilot analysis of new instruments were not conducted; reliability problems in the final report were not acknowledged; and the effects of unreliability in the data on the overall judgment of the programme were not estimated and reported. With the compliance level of five requirements out of ten, the strength of the evaluation was fair.

Systematic Information (AS7): The results of the assessment indicated that protocols for quality control of the evaluation information were established; the evaluation staff who were to adhere to the data protocols and verify data entry were trained; data tables generated from computer output or other means were proofread and verified; storage of the evaluation information was systematized and controlled, data providers verified the data they submitted.

There was no evidence in the evaluation report, which suggested that the accuracy of scoring and coding was systematically checked; that multiple evaluators checked the consistency of their work; that, when feasible, whoever would have access to the evaluation information was identified; furthermore, there was no evidence, which suggested that access to the evaluation information, according to established protocols, was strictly controlled. On this measure, the evaluation complied with six requirements out of ten. In this case the strength of the evaluation was found to be good.

Analysis of Quantitative Information (AS8): The results showed that preliminary exploratory analyses were conducted to ensure the data's correctness and to gain a greater understanding of the data; the evaluation questions and nature of the data procedures were appropriate chosen; each specified procedure on the way its key assumptions were to be met; and visual displays were used to clarify the presentation and interpretation of statistical results.

However, the evaluation case did not comply with six requirements. There was no evidence to suggest that limitations of each analytic procedure, including failure to meet assumptions, were reported; multiple analytic procedures to check on consistency and replicability of findings were employed; variability, as well as central tendencies, were examined; outliers and their correctness were identified, verified and examined; statistical interactions were identified and analysed; or that statistical significance and practical significance were assessed. The strength of this evaluation was found to be fair based on the four requirements that were met out ten.

Analysis of Qualitative Information (AS9): Evidence showed that the evaluation focused on key questions; derived a set of categories, which was sufficient to document; illuminated and responded to the evaluation questions; classified the obtained information into the validated analysis categories; and derived conclusions and recommendations, and demonstrated their meaningfulness.

There was no evidence, based on the contents of the evaluation case, to suggest that boundaries of information to be used were defined nor that information keyed to the important evaluation questions was obtained. Neither was there any evidence regarding the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders; analytic procedures were not employed; no evidence suggested that qualitative information was chosen; no evidence suggested that derived categories for reliability and validity were tested; or that limitations of the referenced information, analyses and inferences were reported. The strength of the evaluation was judged to be fair based on the compliance level of four out of ten requirements.

Justified Conclusions (AS10): The results showed that the evaluation focused conclusions directly on the evaluation questions; the evaluation procedures and findings were accurately reflected; the information that supported each conclusion was cited; and the programme's side effects were identified and reported. However, there was no information found in the contents of the evaluation case that suggested conclusions were limited to the applicable time periods, contexts, purposes and activities, or that plausible alternative explanations of the findings were reported.

There was no explanation in the reports to show why rival explanations were rejected; no evidence of any warning against common misinterpretations; no evidence that the results of a pre-release review of the draft evaluation report were obtained and addressed; or whether or not evaluations limitation were reported. Here the evaluation complied with four requirements out ten and the strength of the evaluation was found to be fair.

Impartial Reporting (AS11): The results showed that appropriate editorial authority was established; the client was engaged to determine steps to ensure fair, impartial reports; and reports were safeguarded from deliberate or inadvertent distortions. However, the contents of the evaluation case did not provide sufficient information to suggest that: right-to-know audiences were determined; appropriate plans for releasing findings to all right-to-know audiences were established and followed; perspectives of all stakeholder groups were reported; alternative plausible conclusions were reported; outside audits of reports were obtained; steps taken to control bias were described; or there was any participation in public presentations of the findings to help guard against and correct distortions by other interested parties. On this standard, the evaluation complied with three requirements out of ten. Therefore, the strength of this evaluation was found to be fair.

Metaevaluation (AS12): The results under this standard showed that responsibility was assigned to someone for documenting and assessing the evaluation process and products. However, insufficient information was obtained in the evaluation case, which provided substantial grounding to demonstrate that the standards used in judging the evaluation were designated or defined; whether or not both formative and summative Metaevaluation was employed; or whether or not the budget was appropriate and sufficient for conducting the Metaevaluation. In addition, there were no evidence suggesting that a full range of information, needed to judge the evaluation against the stipulated standards, was recorded; that an independent Metaevaluation was feasibly contracted; or whether or not audiences that received the Metaevaluation report were determined and recorded.

There was no indication as to whether or not the instrumentation, data collection, data handling, coding and analysis against the relevant standards were evaluated; whether or not the evaluation's involvement of, and communication of, findings to stakeholders against the relevant standards were evaluated; and whether or not a record of all Metaevaluation steps, information and analyses were maintained. With the compliance of only one out ten requirements, the strength of the evaluation on this standard was poor. In summary the overall strength of the evaluation case study on the accuracy standard is fair.

4.2. Summarised Findings on Cross-Case Descriptive Frequency Statistics.

The foregoing section described general findings from document analysis using the PES and associated standards based on each selected evaluation case. The purpose of this chapter is to present descriptive frequency statistics findings across the four case studies. In order to identity the descriptive frequency statistical analysis patterns that emerged, a set of quantitative scales from the quality checklist specified in the methodology chapter were subjected to frequency analysis using the SPSS.

The aim was to show the overall emerging frequency of patterns across the four evaluation cases, as well as the four major categories of evaluation standards composed of Utility (US1-US7) Standards in Table H1 p.228, Feasibility (FS1-FS3) Standards in Table H2 on p.228, Propriety (PS1-PS8) Standards in Table H3 on p.228, and Accuracy (AS1-AS12) Standards in Table H4 on p.228.

In summary, Tables H1 to H4 on p.228 provide summaries of descriptive statistical findings across both the four selected evaluation cases and the four categories of evaluation standards under the PES. This activity helped in synthesising frequency categories of associated evaluation standard based on the PES (for example, US1-US7; FS1-FS3; PS1-PS8 and AS1-AS12) stated above resulting in a summary of results across all four evaluation cases.

The goal was to find out which evaluation standard appears and the most frequently across the selected four cases in order to answer the two research questions (i.e. How effectively did selected evaluations cases conducted under the auspices of the World Bank on poverty-reduction interventions projects funded through the PRSCs in Uganda meet the requirements of a good evaluation based on the PES?).

The mode was, therefore, selected as an appropriate measure of central tendency for purposes of this research. This measure of central tendency appropriately assisted in determining which associated evaluation standard appear the most frequently across the four selected cases based on each major category of evaluation standard under the PES. The results are presented in the following tables:

4.2.1 Cross-case Findings on Accuracy Standards

Table 2: Cross-Case Results on Accuracy Standards (AS1-AS12)

	AS1	AS2	AS3	AS4	AS5	AS6	AS7	AS8	AS9	AS8	AS11	AS12
Valid	4	4	4	4	4	4	4	4	4	4	4	4
Mode	7[a]	5	3	2	3	4	2[a]	4	4	3[a]	3	2

Source: Frequency Statistics findings calculated based on Table D1 on p.205

The results of descriptive frequency statistical analysis across the Accuracy Standards demonstrate that associated standard (AS1) (Programme documentation), with a mode of seven, appears the most frequently across the four cases. This means that seven out of the ten requirements under AS1 attracted very strong compliance across the four evaluation cases.

4.2.2 Cross-Case Results on Feasibility Standards

Table 3: Cross-Case Results on Feasibility Standards (FS1-FS3)

	FS3	FS2	FS1
Valid	4	4	4
Mode	7	4	8

Source: Frequency Statistics findings calculated based on Table B1 on p.200

The results of descriptive frequency statistical analysis across the Accuracy Standards demonstrate that associated standard (FS1) (Practical procedures), with a mode of eight, appears the most frequently across the four cases. This means that eight out of the ten requirements under FS1 attracted very good compliance across the four evaluation cases.

4.2.3 Cross-Case Results on Propriety Standards

Table 4: Cross-Case Findings on Propriety Standards (PS1-PS8)

	PS1	PS2	PS3	PSC4	PS5	PS6	PS7	PS8
Valid	4	4	4	4	4	4	4	4
Mode	7	8	6	5	5	5	4	7

Source: Frequency Statistics findings calculated based on Table C1 on p.202

The results of descriptive frequency statistical analysis across the Propriety Standards demonstrate that associated standard (PS2) (Formal agreements), with a mode of eight, appears the most frequently across the four cases. This means that eight out of the ten requirements under PS2 attracted very good compliance across the four evaluation cases.

4.2.4 Cross-case Results on Utility Standards

Table 5: Cross-case Findings on Utility Standards (US1-US7)

	US1	US2	US3	US4	US5	US6	US7
Valid	4	4	4	4	4	4	4
Mode	2	3	5	4	9	7	4

Source: Frequency Statistics findings calculated based on Table A1 on p.196

The results of descriptive frequency statistical analysis across the Utility Standards demonstrate that associated standard (US5) (Report Clarity), with a mode of nine, appears the most frequently across the four cases. This means that nine out of the ten requirements under US5 attracted excellent compliance across the four evaluation cases.

4.3 Conclusions

General findings from document analysis, without specific reference to specified compulsory standards based on the PES, showed that the strengths of the first and fourth evaluation cases were found to be poor on the basis of the Stakeholder Identification (US1) criterion. On the basis of this criterion, the findings from the assessment of the strengths of selected evaluation cases from document analysis showed that a range of persons involved in, or affected by, the evaluations under consideration were not fully identified and nor were their needs adequately addressed.

The findings from document analysis also revealed that persons who conducted the evaluation were not fully competent to perform the evaluation according to the Evaluator Credibility (US2) Criterion. Deficiencies in the level of trust and competency of evaluators may have undermined the credibility of evaluation outcomes and contributed significantly in exacerbating remarkable inconsistencies in evaluation outcomes in the effectiveness of selected poverty-reduction interventions.

In terms of Defensible Information Sources (AS4) criteria, the study revealed that the strengths of both the first and fourth evaluation cases were also found to be poor. It seemed evident from the findings based on these criteria that the sources of information, which were used in programme evaluation, were not described in enough detail to provide adequate information to support the evaluation process. In addition, evidence from the study showed that the strengths of the second and third evaluation cases were poor on the reliable information (AS6) measure. On the basis of the latter, it was shown in the study that appropriate procedures of gathering information were not well selected and implemented to ensure that the information, which was obtained during the evaluation process, was sufficiently reliable for its intended use.

The study found that quantitative information was not appropriately and systematically analysed so that evaluation questions could be answered effectively. On the basis of the analysis of quantitative information (AS8) measure, it was demonstrated in the study that the strengths of the first and third evaluation cases were also poor. The PES stipulated under Metaevaluation (AS12) that an evaluation itself should be formatively and summatively evaluated against these PES and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strength and weaknesses. It was demonstrated from the findings that the strengths of all four selected evaluation cases were found to be poor on this criterion.

With reference to specified compulsory standards under the PES it was also revealed in the findings that the strength of each of the four selected evaluation cases was not found to be poor, and, therefore none of the selected evaluation cases could be rejected based on compulsory evaluations standards specified under the PES. The evidence generated from the findings revealed that the description and documentation of the programmes attracted the highest level of compliance compared to sources of defensible information; the process of systematically collecting, processing and reporting information; and Metaevaluation, which attracted the lowest levels of compliance across the four selected evaluation cases.

In terms of summarised descriptive frequency statistics, the findings revealed that Report Clarity (US5) was more important across the four selected evaluation cases to the evaluation process than the need to identify a full range of stakeholders [Stakeholder Identification (US1)] and that Practical Procedures (FS1) were more important to the evaluation process than political viability [(Political Viability (FS2)] across all four selected evaluation cases.

Summarised patterns which emerged from descriptive frequency statistical analysis also showed that formal descriptions of sources of information [Programme Documentation (AS1)] were more important to the evaluation process across all four evaluation cases than obtaining information from defensible sources [Defensible Information sources (AS4)]. Summarised descriptive statistics patterns also showed that the acquisition of systematic information [Systematic Information (AS7)] was more important patterns than ensuring the quality of evaluation [Metaevaluation (AS12)] across all four selected evaluation cases.

To recap: Summaries of overall findings in general from document analysis across all four selected evaluation cases are provided in Table F1 on page 225. The findings from document analysis across all four selected evaluation cases on compulsory PES, which will be analysed in detail the next chapter, are presented in Table G1 on page 227. Summaries of findings based on descriptive statistics across all four selected evaluation cases, as well as all four categories of evaluation standards under the PES, are provided under Tables H1, H2, H3 and H4 on page 228.The next chapter analyses the implications of these findings.

CHAPTER 5: ANALYSIS OF THE FINDINGS

5.0 Introduction

The problem this research sought to solve was that while it was evident that selected poverty-reduction interventions funded by the World Bank through the PRSCs in Uganda had been systematically evaluated in order to measure their effectiveness, there was overall evidence of remarkable inconsistencies in the evaluation outcomes of the effectiveness of these selected poverty reduction interventions. Using the mixed research methods approach, the main purpose of this study was to clarify and resolve the evident causes of such remarkable inconsistencies in evaluation outcomes of the effectiveness of selected poverty reduction interventions.

The preceding chapter presented a description of the findings, based on the four selected evaluations cases in Uganda's case study. The overall purpose of this chapter is to analyse these findings in order to describe emerging evidence of knowledge cues. This chapter is divided into five main sections: the first section analyses general findings from document analysis across all selected evaluation cases, so that, general lessons evolving from the findings on the extent to which each evaluation case complied with prescribed evaluation standards under the PES can be identified. The second section analyses specific findings based on compulsory associated standards under the PES in order to determine the extent to which each evaluation case met the requirements of a good evaluation based on the compulsory associated prescribed under PES.

The third section analyses the findings from summarised descriptive frequency statistics based on each category of standards as well as across the four selected evaluation cases. The aim of the third section is to determine overall emerging frequency of patterns and discover which associated standard appeared the most frequently across the selected four evaluation cases, in order to determine the associated standard, which attracted the highest compliance levels across these four cases.

The fourth section clarifies emerging aspects and evidence of new knowledge from the analysis of the findings. The aim of this section is to tease out glaring issues emerging from the findings which will constitute the basis for formulating statements on contributions to knowledge. The fifth section sketches out conclusions to the analysis.

5.1 Summarised Analysis of General Findings from Document Analysis

The strengths of the first and fourth evaluation cases were found to be poor on the basis of the Stakeholder Identification (US1) criterion. On the basis of this criterion, the findings from the assessment of the strengths of selected evaluation cases from document analysis showed that a range of persons involved in, or affected by, the evaluations under consideration were not fully identified nor were their needs adequately addressed. The findings from the document analysis also revealed that persons conducting the evaluation were not fully competent to perform the evaluation according to the Evaluator Credibility (US2) Criterion. Deficiencies in the level of trust and competency of evaluators may have undermined the credibility of evaluation outcomes and the strength of the fourth evaluation case, which was also judged to be poor.

In terms of Defensible Information Sources (AS4) criterion, the findings from document analysis revealed that the strengths of both the first and four evaluation cases were also found to be poor. It seemed evident from the findings based on AS4 criteria that the sources of information, which were used in a programme evaluation, were not described in enough detail to provide adequate information to support the evaluation process. In addition, evidence showed that the strengths of the second and third evaluation cases were poor on the Reliable Information (AS6) measure. On the basis of this measure, the findings showed that appropriate procedures of gathering information were not well selected or implemented to ensure that the information, which was obtained during the evaluation process, was sufficiently reliable for the intended use.

In terms of the analysis of quantitative information evaluation, the findings showed that quantitative information was not appropriately and systematically analysed so that evaluation questions could effectively be answered. On the basis of the Analysis of the Quantitative Information (AS8) measure, the findings demonstrated that the strengths of the first and third evaluation cases were also poor.

The PES stipulated under Metaevaluation (AS12) that an evaluation itself should be formatively and summatively evaluated against these PES and other pertinent standards, so that its conduct can be appropriately guided, and on completion, stakeholders can closely examine its strengths and weaknesses. The findings demonstrated that the strengths of all four selected evaluation cases were found to be poor on this criterion. (See also a summary of these findings in Table F1, p.225).

5.2 Analysis of the Findings from Compulsory Evaluation Standards

5.2.1 Impartial Reporting (AS11): Impartial reporting can be defined as a situation where one operates in an objective manner devoid of one’s own personal views. The findings across all four selected evaluation cases showed that appropriate editorial authority was established and that the right-to-know audiences were determined. The findings also demonstrated that appropriate plans for releasing findings to all right-to-know audiences were established and followed.

Whilst it was specified in the literature on Metaevaluation according to Sanders (1994) and Stufflebeam (2005) that reporting procedures should guard against distortion caused by personal feelings and bias of any party to the evaluation in order that evaluation reports can reflect the evaluation findings fairly, the Metaevaluation findings across the four selected evaluation cases demonstrated profoundly low compliance levels. The findings revealed that all four evaluation cases did not comply with seven out of ten requirements under this measure. These findings underscored the extent to which a lack of significant attention was paid to the role played by impartial reporting in fostering the legitimacy of evaluations during the evaluation process.

The findings in this case showed that there was no evidence, which indicated that the process of evaluating selected poverty-reduction interventions was protected from deliberate or inadvertent distortions, let alone reporting the perspectives of all stakeholder groups. The lack of substantial commitment in safeguarding the evaluation process of these selected poverty-reduction interventions against unintended distortions caused by either personal feelings or bias of any party had the potential of precipitating inappropriate interferences, which in turn had rather debilitating consequences on the quality of the evaluation outcomes.

The findings across the four selected cases also showed that all four evaluations failed to provide sufficient information to suggest that the client was engaged in determining steps to ensure fair and impartial reports. These findings uncovered a testimony of the extent to which reporting of evaluation of such critical interventions was undermined, it being intended to provide greater clarity on the effectiveness of the intervention; hopefully to inform policy; and the kinds of policy change and possible influence on holistic change in the programmatic architecture of the intervention and/or change of some aspects of the intervention or an entire intervention.

Since all four selected evaluation cases complied with three out of ten requirements, the strengths of all four evaluation cases were found to be fair. Therefore, in terms of the recommendations, based on compulsory associated evaluation standards under the PES, none of the four selected evaluation cases could be rejected.

5.2.2 Service Orientation (PS1).

The findings on this compulsory measure revealed that the needs of the programme's customers were assessed; excellent service was promoted; the evaluation's service orientation was made clear to stakeholders; and interim feedback for programme improvement was given. Despite identifying the strengths on which to build within the programme, and correcting its weaknesses, the findings did, however, demonstrate that the evaluation process did not help to ensure that the full range of rightful programme beneficiaries was served, and that the outcomes of the programme were not assessed against targeted customers' assessed needs.

In comparison to other compulsory associated standards, there were notable contradictions that significantly emerged. With specific reference to the findings under the Impartial Reporting (AS11) Associated evaluation Standard in Section 5.1.1, whilst it was shown that the right-to-know audiences were determined, nevertheless, under this section, evidence suggests that the evaluation process did not help to ensure that the full range of rightful programme beneficiaries was served, and that the outcomes of the programme were not assessed against targeted customers' assessed needs. In addition, the findings also revealed that harmful practices were not exposed, coupled with the fact that all right-to-know audiences were not informed of the programme's positive and negative outcomes.

According to Sanders (1994), this compulsory measure of the worth and merit of an evaluation stipulates that evaluations should be designed to assist organisations to address and to serve the needs of the full range of targeted participants effectively. Evidence from the findings demonstrated that the first evaluation case complied with six out of ten requirements, and the strength of the evaluation case was, therefore, judged to be good. The findings on the other three evaluation cases under consideration had very good levels of compliance with scores of eight requirements out of the maximum levels of ten.

Nevertheless, the fact that the needs of the full range of targeted participants were not effectively served and addressed was of great concern to the researcher. These results are by implication reflective of analogous opinion which was highlighted by the World Bank in its evaluation report of the overall strategic Africa Action Plan (AAP). In its 2010 AAP report, the World Bank reiterated that "the process of design and evaluation was largely—top-down with little consultation with (and hence ownership by) country clients and stakeholders" (World Bank 2010, p.20).

With this evidence, it became clear that the lack of participation by a full range of rightful stakeholders in both designing and evaluating these poverty-reduction interventions raises serious questions about the legitimacy of the evaluation results. It can be urged here that the inclusion of a full range of rightful stakeholders in designing and evaluating these poverty-reduction interventions would have enhanced both the usefulness and legitimacy of evaluation results of selected poverty reduction interventions. The inclusion of a full range of rightful stakeholder heightens the importance of broad ownership of the outcomes of the decision-making processes if such knowledge is produced jointly and used by a wide range of stakeholders, including persons and the communities with an interest in the outcomes of the intervention(s).

Therefore, it can be adjudged that the lack of participation by a full range of rightful stakeholders in designing programme interventions, as well as the evaluation of the outcomes of these interventions, may have also undermined the importance of enhancing participatory democracy in development processes. Nevertheless, according to the PES, the findings on the four evaluation cases provided sufficient evidence to determine that none of the four selected evaluation cases could be rejected based on the Service Orientation (PS1) as a compulsory standard. (See also a summary of these findings in Table G1, p.227).

5.2.3 Valid Information (A5)

Recommendations on compulsory associated evaluation standards under the PES state that the information gathering procedures should be chosen or developed, and then implemented, so that they will ensure that the interpretation, which is arrived at, is valid for the intended use (Sanders, 1994 and Stufflebeam, 2005). The findings suggested that whilst the evaluation process focused the evaluation activities on key questions and appropriately employed multiple measures to address each question while training and calibrating data collectors, a detailed description of the constructs and behaviours was not provided regarding the information that was acquired.

The findings also showed that the type of information each employed procedure acquired was not assessed or reported. Whether or not the collection conditions and process were documented and reported is not clear from the findings. The findings also do not provide sufficient information to determine how information from each procedure was scored, analysed, interpreted and documented.

There was also insufficient information in the findings from both the first and fourth selected evaluation cases to provide the basis for judging whether or not inferences, singly and in combination, were reported and justified. The examination of the findings for both the first and fourth evaluation cases showed that the comprehensiveness of a set of information provided by the procedures in relation to the information needed to answer the set of evaluation questions was not assessed and reported.

Based on the findings from the first and fourth evaluation cases, it is clear that meaningful categories of information, which required identifying regular and recurrent themes in information collected through the evaluation process using qualitative assessment procedures, could not be established.

Nevertheless, the findings from the second and third selected evaluation cases significantly revealed similar patterns compared to the patterns identified in the findings of the first and fourth cases, which have been considered in the preceding section. In the case of the second and third selected evaluation cases, the findings revealed, for example, that multiple measures to address each question were not employed appropriately to assess and report on the type of information each employed procedure acquired.

The findings across the second and third selected evaluation cases revealed that data collection conditions and processes were not documented and reported, and inferences, whether singly or in combination, were not reported and justified. While the findings across the four evaluation cases revealed evidence of inconsistency, it can be added that these inconsistencies are clearly in contrast with the basic requirements of the evaluation profession when conducting evaluations, in order to assess the effectiveness of an intervention. In seeking to determine the effectiveness of an intervention, evaluations are conducted, so that institutions of accountability have access to reliable and consistent information on the performance of the interventions.

The generation of consistent evaluation information is, therefore, critical for improving both accountability and decision-making. In this way, decision-makers can also feel compelled to respond with consistency to the demands of various stakeholders for action, and to continue to provide feedback on the effectiveness of an intervention. It can also be added that institutions of accountability get reliable and evident motivation to take action, based on such information generated through evaluation.

Whilst categorical evidence of inconsistencies between the two sets of evaluation cases are evident, the strengths of the first and fourth cases and the strengths of the second and fourth cases, respectively, were adjudged to be fair and good. However, across the four evaluations, insufficient factual evidence exists to reject all four cases based on the recommendations under the PES with respect to the Valid Information (A5) criterion as a compulsory associated evaluation standard.

5.2.4 Justified Conclusions (A10)

The Joint Committee on Program Evaluation Standards explained that the justification of conclusions is about making claims regarding the programme that are warranted on the basis of data that have been compared against pertinent and defensible ideas of merit, value or significance against specified standards of values. The Joint Committee also contends that conclusions are justified when they are linked to the evidence gathered and consistent with agreed values or standards of stakeholders (The Joint Committee, 1994).

The findings revealed that the strengths of all four evaluation cases were found to be fair under this measure. While the strength of each case was found not to be poor, the findings did reveal, however, evidence of significant variations therein at the level of an individual selected evaluation case. The need to analyse these significant variations was necessary in order to tease out elements that caused such variations. An examination of individual requirements revealed that the first and the second evaluation cases complied with four out of the ten prescribed requirements, whereas the third and fourth evaluation cases complied with five requirements out ten prescribed requirements.

The findings on both the first and second selected evaluation cases revealed that the conclusions focused directly on the evaluation questions, in which the evaluation procedures and findings were accurately reflected, coupled with a focus on the applicable time periods, contexts, purposes and activities of evaluations. In addition, the third and fourth evaluation cases cited information that supported each conclusion.

Furthermore, the findings revealed that the programme's side effects were not identified and reported, and plausible alternative explanations of the findings were not reported across the four selected evaluation cases. The findings also demonstrated the lack of explanation as to why rival explanations were rejected across the four selected evaluation cases.

There were also no indications, in all four selected cases, of any warning against common misinterpretations. In all four cases, it is not very clear from the findings that pre-release review of the draft evaluation report was obtained, and whether or not the contents addressed and reported the limitations of the evaluations. In comparison, the first and second evaluation cases met three out of the ten prescribed requirements, whereas both the third and fourth evaluation cases complied with four out of ten requirements.

Sanders (1994) required, in the literature review on Metaevaluation (in Chapter 2, Section 3.2.6) that conclusions reached in an evaluation should be explicitly justified so that stakeholders can assess them. It is very clear from the findings and the subsequent analysis that, given such low levels of compliance across the four selected evaluation cases, the findings are ineffective in providing substantial factual evidence to submit that conclusions were substantially justified. Notwithstanding these inadequacies, no grounds were provided, based on factual information provided in the findings, to fail the four evaluation cases under this Justified Conclusions (A10) measure based on the PES.

5.3 Analysis of findings based on Descriptive Frequency Statistics

This section analyses the summaries of findings from descriptive frequency statistics across the four selected evaluation cases. The aim is to show overall frequency of patterns across the four cases, as well as across all associated standards, and to identify the associated standard that appeared the most frequently across the selected four cases, in order to determine the associated standard, which attracted the highest compliance levels across the four cases. (See summary in the Tables H1, H2, H3 and H4 on p.228).

5.3.1 Cross-Case Analysis on Utility Standards.

Compliance with Utility Standards was meant to ensure that the information generated from evaluations served the needs of intended users by paying attention to stakeholder identification, evaluator credibility, information scope and selection, values identification, report clarity, report timeliness and dissemination, and report impact (Stufflebeam (2001, 2005), Rogers (1992), Weiss (1983), Patton (1978, 1986, and 1997). The findings demonstrated that, compared to Stakeholder Identification (US1), which surprisingly attracted the lowest compliance level across the four selected evaluation cases, Report Clarity (US5) attracted the highest level of compliance across the four cases.

These findings, therefore, suggested that ensuring that evaluation reports clearly described the programme being evaluated, including its context; and the purposes and findings of the evaluation, so that essential information being provided and easily understood was of greater importance than ensuring that persons who involved in, or affected by, the evaluation needed to be identified, thus ensuring that their needs could be addressed throughout the evaluation process.

In summary, the analysis of the findings from descriptive frequency statistics in the preceding sections across the four selected evaluation cases demonstrated that formal agreements attracted high levels of compliance compared to the requirements to avoid conflict of interest (Conflict of Interest, PS7). The analysis also showed that clear and accurate description and documentation of the programme to be evaluated seemed to be more important than sources of defensible information; process of systematically collecting; processing and reporting information; and ensuring the quality of evaluations. Evidence also suggested that complying with practical procedures in the evaluation process was of greater importance. Finally, stakeholder identification, surprisingly, attracted the lowest compliance level of two out of ten requirements across all four selected evaluation cases. The next section scrutinises emerging aspects from the analysis (see also summary in the Table H1 on p.228).

5.3.2 Cross-Case Analysis on Feasibility Standards

The findings revealed that Practical Procedures (FS1), such as tailoring methods and instruments to information requirements; minimising disruption through the evaluation process; minimising the data burden; appointing competent staff; training staff; choosing procedures that the staff were qualified to carry out; choosing procedures in light of known constraints; making a realistic schedule; engaging locals to help conduct the evaluation; and as appropriate, making evaluation procedures part of the routine achieved compliance levels of eight out of ten requirements with a mode of eight across the four evaluation cases.

The findings also revealed that, compared to the requirements (such as: anticipating differing positions of different interest groups; averting or counteracting attempts to bias or misapply the findings; fostering cooperation; involving stakeholders throughout the evaluation; agreeing on editorial and dissemination authority; issuing interim reports; reporting divergent views; reporting to right-to-know audiences; employing a firm public contract; and, terminating any corrupted evaluation), which are associated with political viability criterion under feasibility standards, each of the four selected evaluation cases complied with eight requirements out of ten requirements which are associated with practical evaluation procedures. In the case of the findings on feasibility standards, evidence suggested that complying with practical procedures in the evaluation process was of greater importance to evaluators than complying with the need to anticipate differing positions of various interest groups in order to obtain their cooperation, so that possible attempts by any group to either curtail evaluation operations or misapply the results can be avoided (see also summary in the Table H2 on p.228).

5.3.3 Cross-Case Analysis on Propriety Standards

The purpose of propriety standards is to ensure that an evaluation is conducted legally, ethically and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results (Sanders, 1994). The Propriety Standards comprised eight criteria (PS1-PS8). The findings showed that all four cases complied with formal agreements (PS2) compared to the other seven associated criteria, such as Service orientation, rights of human subjects, human interactions, complete and fair assessment, and disclosure of findings, conflict of interest and fiscal responsibility.

However, it is surprising to discover from these findings that while eight requirements associated with formal agreements attracted high levels of compliance, conflict of interest attracted the lowest level of compliance at four out of the ten requirements. These findings demonstrate that agreements in writing on the obligations of formal parties to an evaluation took precedence over the obligation to deal openly and honestly with any conflict of interest, which could have potentially compromised the evaluation processes and results (see also summary in the Table H3 on p.228).

5.3.4 Cross-Case Analysis on Accuracy Standards

The descriptive frequency statistics findings on this measure established that among the twelve associated under accuracy standards, “Programme documentation (AS1)”, with a mode of seven, appeared the most often throughout the results of frequency analysis across the four selected evaluation cases. This means that all the four cases complied with seven out of the ten requirements associated with this criterion under the Accuracy Standards. Compared to other criteria associated with accuracy standards, the description and documentation process of the programme attracted the highest level of compliance. Sources of defensible information, the process of systematically collecting, processing and reporting information and Metaevaluation attracted the lowest level of compliance with only two out ten requirements across all four cases.

The findings, in summary, revealed that clear and accurate description and documentation of the programme to be evaluated seemed to be more important to evaluators than ensuring that the process of evaluation and the evaluation itself was formatively and summatively evaluated against pertinent standards. The importance of formative and summative evaluation is that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strength and weaknesses (see also summary in the Table H4 on p.228).

5.4 Clarification of Emerging Aspects and evidence of new Knowledge from the Analysis.

The objective of this section is to scrutinise evidence of emerging evaluation aspects from the research, which may have caused such remarkable inconsistencies in evaluation outcomes of the effectiveness of poverty-reduction interventions of the World Bank, funded through its PRSCs in Uganda so that defensible grounds can be provided to answer the following research question: How effectively did evaluations, conducted under the auspices of the World Bank on poverty-reduction interventions funded through the PRSCs, meet the requirements of a good evaluation?

The evidence from the analysis of findings revealed that the strength of none of the four selected evaluation cases was found to be poor and therefore none of the selected evaluation cases could be rejected, based on compulsory associated evaluations standards specified under the PES. The evidence generated from the analysis included clear proof drawn from the analysis of the findings, which demonstrated that the description and documentation of the programmes attracted the highest level of compliance, compared to: sources of defensible information; the process of systematically collecting, processing and reporting information; and Metaevaluation, which attracted the lowest levels of compliance across the four cases.

The knowledge cue that emerged from the analysis was that clear and accurate description and documentation of the programme to be evaluated were more important to the evaluators than ensuring that the process of evaluation, and the evaluation itself, was formatively and summatively evaluated against pertinent standards so that its conduct is appropriately guided, and, on completion, stakeholders could closely examine its strengths and weaknesses.

In addition, the process of analysis also uncovered evidence, which suggested that complying with practical procedures in the evaluation process was more important than complying with the need to anticipate differing positions of various interest groups, in order to obtain their cooperation so that possible attempts by any group to either curtail evaluation operations or misapply the results could be avoided.

The analysis also revealed sufficient information showing that agreements in writing on obligations of formal parties to an evaluation took precedence over the obligation to deal openly and honestly with conflict of interest, which could have, potentially, compromised the evaluation processes and results. Furthermore, the evidence from the analysis illustrated that ensuring that evaluation reports clearly described the programme being evaluated, including its context, the purposes and findings of the evaluation, so that essential information is provided and easily understood, was of greater importance to the evaluation process, than ensuring that persons who are involved in, or affected by, the evaluation needed to be identified, so that their needs could be addressed throughout the evaluation process.

Nevertheless, whilst the analysis of the findings provided significant new strands of knowledge cues and lessons regarding the eminent causes of evident inconsistencies in evaluation outcomes of the effectiveness of selected poverty reduction interventions of the World Bank funded through its poverty reduction support Credits (PRSCs), there were variations in the extent to which each of the four selected evaluation cases complied with the prescribed requirements.

In general, evidence from the analysis of findings from document analysis demonstrated that none of the selected evaluation cases failed to meet the prescribed compulsory criteria based on the PES. Furthermore, the clarification of emerging knowledge cues from the analysis of the findings clearly revealed that agreements in writing on obligations of formal parties to an evaluation was more important than the obligation to deal openly and honestly with conflict of interest, which could have potentially compromised the evaluation processes and results.

The absence of the evidence that conflict of interest was openly dealt with in particular, may have exposed the evaluation process to inappropriate influences, which may have seriously undermined the ability of the evaluation process to make unbiased constructions on the effectiveness of these poverty-reductions interventions. The minimal attention paid to the avoidance of conflict of interest may have undermined the evaluation process from exercising due judgement and decision-making tasks on the effectiveness of the evaluand, thus stifling the commitment to both accountability and transparency responsibilities.

In answering the research question [How effectively did evaluations, conducted under the auspices of the World Bank on poverty-reduction interventions funded through the PRSCs, meet the requirements of a good evaluation?]: The evidence suggests that none of the selected evaluation cases could be rejected based on compulsory associated evaluations standards prescribed in evaluation literature under the PES.

However, it is evidently clear from the analysis that the four selected evaluations cases did better on bureaucratic criteria and worse on accountability and transparency procedures. Firstly, the analysis of the findings glaringly revealed that a full range of persons involved in, or affected by, the evaluation was not adequately identified. The involvement of a full range of stakeholders in, not only designing social development interventions, but also evaluating these social interventions, is essential in promoting the principles of accountability and transparency. In this case, the analysis demonstrated that compliance with Report Clarity (US5) criteria was more important across the four selected evaluation cases than the need to identify a full range of stakeholders, based on the Stakeholder Identification (US1) criteria.

The lack of significant attention paid to the importance of the identification and engagement of a full range of stakeholders that had an interest in, or were affected by poverty reduction interventions throughout the evaluation process, is of great concern in terms of the evaluation profession and practice. In situating metaevaluation in evaluative inquiry, Orr, Cameron and Day (1991) noted in Chapter Two (section 2.1) that although there had been a proliferation in evaluation models, a theoretical framework, which addressed the 'why' (why an evaluation is conducted) from the point of view of the evaluation practitioner was missing where such a question has significant implications on the manner in which evaluation is conducted.

Owen and Rogers (1999) developed a meta-model consisting of five categories of evaluative enquiry to address the 'why' question in evaluation practice from the perspective of an evaluation practitioner. One of the five meta-models was the 'interactive' evaluation meta-model, which is based on the assumption that those with direct vested interest in the intervention should also control the evaluation of the intervention (Owen and Rogers 1999). While the analysis of the findings across the four evaluation cases revealed evidence of inconsistency with the requirements of the practice of good evaluations, it can also be added that these inconsistencies were clearly not consistent with the basic requirements, which undergird the practice of evaluation as a profession.

In seeking to determine the effectiveness of social development interventions, evaluations are conducted so that parties of accountability and transparency are fully engaged, not only in designing and evaluating the interventions, but also have access to reliable and consistent information on the performance of social development interventions. Alkin (1969, 2012) also argued in the literature review that the evolution of the theory for good evaluation was driven mainly by the quest to promote accountability, transparency and social enquiry.

In the main, these postulations on the quest to promote accountability and social enquiry implicitly call for the democratisation of the process of development in order to sustain the effectiveness of development interventions. The democratisation of the process of development and the sustainability of the effectiveness of development interventions cannot be successfully implemented without adequate identification and engagement of a full range of stakeholders.

The findings were dissimilar across the four selected evaluations cases on the basis of the democratisation of the process of development within the context of the Stakeholder Identification (US1) criteria. On the basis of these criteria, the first and fourth evaluation cases complied with only two out of ten requirements. The findings also revealed that both the second and third evaluation cases complied with six requirements out ten requirements. In the case of the first and fourth evaluation cases, the findings demonstrated that the evaluation client was clearly identified and potential stakeholders were consulted, in order to identify their information needs.

However, the findings indicated that both these latter cases failed to comply with eight requirements. In both the first and fourth cases, leadership figures were not engaged to identify other stakeholders. In addition, stakeholders were not used to identify other stakeholders. The findings also showed that stakeholders were not ranked according to their relative importance with the client. The findings also revealed that no arrangement was made to involve stakeholders throughout the evaluation process.

In addition, the evaluation process was not kept open to serve newly-identified stakeholders; also the evaluation did not address the needs of stakeholders; and the findings also showed that neither appropriate ranges of individual stakeholders and organisations were served. In the case of the second and third evaluation cases, where both cases complied with six out ten requirements, the evaluation client was not used to identify stakeholders.

Unlike the case of the first and fourth evaluation cases, the second and third evaluation cases used stakeholders to identify other stakeholders. Leadership figures were also engaged to identify other stakeholders and the evaluation process was kept open to serve newly-identified stakeholders in the case of the second and third evaluation cases.

Furthermore, in both the second and third evaluation cases, the needs of stakeholders were addressed and appropriate ranges of individual stakeholders, as well as organisations, were served. The outcomes of the analysis of the evidence related to the first and fourth evaluation cases emphasised the centrality of the role played by stakeholders in, not only enhancing the effectiveness of development interventions, but also in promoting good governance practices, such as democratisation and accountability, both of which, if used properly, constrain the exercise of power by holding decision-makers accountable.

Secondly, the analysis revealed that persons who conducted the evaluation were not fully competent to perform the evaluation according to the Evaluator Credibility (US2) criteria, and, by implication, deficiencies in the level of trust and competency of evaluators could have undermined the credibility of the evaluation outcomes.

Thirdly, in terms of Defensible Information Sources (AS4) criteria, the analysis of the findings from document analysis revealed that the sources of information, which were used in a programme evaluation, were not described in enough detail to provide adequate information to support the evaluation process.

Fourthly, in terms of the Reliable Information (AS6) measure, the analysis of the findings showed that appropriate procedures of gathering information were not well selected enough and implemented to ensure that the information, which was obtained during the evaluation process, was sufficiently reliable for the intended use.

Fifthly, the analysis of the extent to which quantitative information was systematically analysed showed that quantitative information was not appropriately or systematically analysed in order to answer evaluation questions effectively (Analysis of Quantitative Information AS8).

Finally, the findings from the analysis revealed that selected evaluations were not, themselves, formatively or summatively evaluated against these PES and other pertinent standards so that stakeholders could closely examine their strengths and weaknesses (Metaevaluation AS12).

Furthermore, such evidence of robust performance on bureaucratic criteria compared to the ineffective compliance with accountability and transparency procedures contradicts the raison d'etre for the intervention of the World Bank in Uganda. The evidence of robust performance on bureaucratic criteria compared to the ineffective compliance to accountability and transparency requirements, which the World Bank sought to establish by trimming bloated bureaucracies in Uganda can be subsumed in the argument that failures in achieving accountability and transparency reforms in Uganda by the World Bank can be attributed to inappropriate development policy assumptions and prescriptions, which were pursued by the World Bank to support Uganda’s reform process.

It was learnt in the literature review in Chapter 2 (section 2.2.2) that the World Bank directed its efforts towards trimming bloated state bureaucracy in African countries, including Uganda, into a leaner and more efficient government, necessary to spur social development (Toye 1991, 1993). The worse performance on accountability and transparency procedures compared to the better performance on bureaucratic criteria clearly provided evidence of the ineffective nature of development policy assumptions and prescriptions pursued by the World Bank in Uganda's case study and raises a lot of concern.

Firstly, it can be argued that the policy prescriptions, which were created by the World Bank to transform Uganda's defective political and economic systems, undoubtedly reinforced the extensive interventionist influence of the state in development processes.

Whilst the rationale for the involvement of the World Bank in Uganda was to restructure state controls resulting in trimmed and less-bloated state bureaucracy, Barkan (2011) recently showed that Uganda's system of governance is "in essence an authoritarian patronage-based regime" (p.2). In addition, Mwenda and Tangri (2005) had argued earlier that the relationship between the donors and the Ugandan government has reproduced state patronage and that this relationship has "provided the government with large amounts of financial assistance to sustain the patronage system of governance" (p.449).

Furthermore, the architecture of governance in Uganda, under an authoritarian patronage-based regime, is characterised by what Diamond (2008) described as an "insidiously overpowering presidency punctuated by an overpowering presidency"(p.139), which is sculptured around patronage, grounded in ethnic bonding and punctuated by a hierarchical network of patron-client form of state organisation(Mwenda and Tangri 2005). Therefore, the persistent sustenance of such a patronage system of governance in Uganda had debilitating effects on efforts to eradicate grinding poverty.

Secondly, it is evident that whilst working with the Ugandan government, the World Bank implicitly operated under the assumption that development takes place in a vacuum, and that the government cares about its people. It can be argued within the context of this notion that structural adjustment programmes, which the Ugandan government undertook, were sponsored under the assumption that defective governance systems do not affect equitable development in Uganda.

However, Uganda's authoritarian patronage-based regime, with such an insidiously overpowering presidency that is sculptured around patronage, grounded in ethnic bonding and punctuated by a hierarchical network of patron-client system of governance (Mwenda and Tangri 2005; Diamond, 2008; Barkan; 2011), fundamentally undermines broad equitable social development efforts. Instead, such a system inevitably supports selective ethnic-based development, which is based on ethnic bonding.

In addition, the assumption that development takes place in a vacuum and that the government cares about its people clearly ignores the evident reality that government, within the context of an authoritarian patronage-based regime, inescapably transmogrifies into a criminal enterprise. The effects of authoritarian patronage-based regimes in African countries are well documented.

Ackoff (1977); Ayittey (1991) and Ndikumana and Boy (2011) state that authoritarian patronage-based African governments characteristically dominate the economy and transfer resources to the state through excessive state controls, and that such resources are used for the benefit of a tiny ruling elite, turning the public sector into their personalised employment and money-manufacturing mills.

Specifically, Ayittey (1991) highlighted the fact that the well-carved bureaucratic machinery with payrolls are stuffed with government or party supporters breeding increased opportunities for illicit enrichment by the elite, associated with the authoritarian patronage-based regime.

Therefore, it is inevitable that the Ugandan government could not be interested in genuine political and economic reforms, which explain better performance on the bureaucratic criteria compared to accountability and transparency-based criteria of selected evaluation cases. The efficient implementation of political and economic reforms would strip members of the authoritarian patronage-based regime of their power to allocate foreign financial resources to themselves, their cronies and their relatives at the expense of broader social development imperatives of the poverty stricken masses.

5.5 Conclusions

The analysis of the findings showed that a range of persons involved in, or affected by, the evaluations under consideration were not fully identified and their needs were not adequately addressed. The analysis also showed that deficiencies in the level of trust and competency of evaluators undermined the credibility of evaluation outcomes and that sources of information, which were used in a programme evaluation, were not described in enough detail to support the evaluation process.

The analysis of findings from document analysis showed that appropriate procedures of gathering information were not well-selected or implemented to ensure that the information, which was obtained during the evaluation process, was sufficiently reliable for the intended use, and that quantitative information was not appropriately and systematically analysed so that evaluation questions could be answered effectively. The analysis revealed that selected evaluation themes were not formatively and summatively evaluated against pertinent standards so that their conduct could be guided appropriately and, on completion, stakeholders could closely examine their strengths and weaknesses.

While the analysis provided significant lessons, there were variations in the extent to which each of the four selected evaluation cases complied with the prescribed requirements. Notwithstanding these variations in the extent to which each of the four selected evaluation cases complied with the specified compulsory associated standards of a good evaluation under the PES, the analysis demonstrated that the strength and quality of each of the four selected evaluation cases was not found to be poor.

However, despite this conclusion, the evidence from the analysis nevertheless revealed that the four selected evaluations cases performed better on bureaucratic criteria and worse on accountability and transparency procedures. The failure to promote accountability and transparency through trimming bloated bureaucracies in Uganda can be attributed to both inappropriate development policy assumptions and prescriptions, which were pursued by the World Bank in support Uganda's reform process. The next chapter concludes the thesis, outlining the contribution to knowledge and suggesting areas for further Metaevaluation studies.

CHAPTER 6: CONCLUSIONS, CONTRIBUTION TO KNOWLEDGE AND AREAS FOR FURTHER RESEARCH

6.1 Conclusions

Situated in evaluative enquiry, Metaevaluation was defined as evaluation of evaluations. Metaevaluation is a unique and specialised form of evaluation, which is grounded in the theory for good programme evaluation and idealised within the context of solving social problems faced by society. Metaevaluation helps in judging the strengths and weaknesses of primary evaluations so that interested parties can be able to judge the reliability and credibility of evaluations and the evidence used to make decisions.

Metaevaluation had been utilised in different contexts in some developed countries to assess the quality of evaluations of development interventions. The application of Metaevaluation to assess the quality of evaluations of development interventions had, however, remained scanty in developing country contexts, particularly in Uganda, until I embarked on this study.

Uganda, after many years of conflict, which ended in 1986, adopted governance policies directed at stabilisation, rehabilitation and recovery that endeared the World Bank to provide its government with funding through an assortment of PRSCs to support the country's poverty-reduction efforts.

In practice, development funders of poverty-reduction interventions are mainly interested in steady consistency in the effectiveness of their funding support. A scrutiny of evaluation outcomes of the effectiveness of the four selected World Bank poverty-reduction interventions, funded through its PRSCs in Uganda, however, revealed evidence of remarkable inconsistencies in evaluation outcomes in the effectiveness of selected poverty-reduction interventions of the World Bank in Uganda, which were funded at successive annual intervals from 2001 to 2005. These inconsistencies in evaluation outcomes of predominantly identical poverty reduction interventions which were implemented repeatedly over the five years raised serious questions about the legitimacy and quality of the evaluations of these selected poverty-reduction interventions.

The questions about the legitimacy and quality of selected evaluations, as well as the demand by development funders for steady consistency in the effectiveness of their funding support, reinforced the importance of evaluating the quality of evaluations of these poverty-reduction interventions in order to identify and resolve the causes of such inconsistencies.

The methodology for evaluating the quality of four selected evaluations of poverty reduction interventions funded by the World Bank through the PRSCs was based on the mixed methods approach grounded in constructivist evaluation within the context of the constructivist paradigm.

Constructivist evaluation propagates the construction of both merit judgments on the intrinsic quality of the evaluand (intervention/project) and worth judgments on the extrinsic usefulness of the evaluand. The constructivist philosophical paradigm which undergirded this study is based on three fundamental assumptions of knowledge which include ontological assumptions of relativism, epistemological assumptions of subjectivism and methodological assumptions of hermeneutic-dialecticism.

The mixed methods approach was composed of a case study approach, document analysis and descriptive frequency statistics. The selected mixed methods assisted in identifying evident causes of such remarkable inconsistencies in evaluation outcomes of selected evaluation cases, as well as to provide answers to the research question.

In assisting to identify evident causes of such remarkable inconsistencies, selected mixed methods provided the appropriate leverage for triangulating multiple perspectives and information sources in order to generate credible evidence.

Triangulation of different selected methods also protected the study from the generation of evidence based on what could be construed as pieces of evidence based on a single method. The use of mixed methods also protected evaluation of the quality of selected evaluations from probable bias associated with the use of a single method to generate empirical evidence.

On the basis of selected evaluation cases, a case study approach was appropriate in addressing accuracy issues by employing and triangulating multiple perspectives and information sources. The closeness of selected evaluation cases to real–life situations provided a wealth of details and useful strands of knowledge. In addition, the use of multiple study cases enhanced external validity and helped to guard against observer bias.

In particular, the use of document analysis involved identifying, choosing, valuing and integrating data obtained from documents, using the four selected evaluation cases. The prescribed quantitative scales, based on the quality checklist, prescribed under the PES, were used for scoring the quality of evaluations. The judgments about the adequacy and strength of each selected evaluation case in meeting the requirements of a good evaluation, based on the PES were made, based on a set of prescribed quantitative scores.

Each of the four selected evaluation cases in Uganda's case study was subjected to a checklist of 300 requirements prescribed under the PES. There were 1,200 checklists for all the four selected evaluation cases. Four compulsory associated evaluation standards specified under the PES, which were used were: Service Orientation (P1), Valid Information (A5), Justified conclusions (A10) or Impartial Reporting (A11) on any the four compulsory associated standards. Judgments about the adequacy of the selected evaluation cases in meeting the standards of a good evaluation under the PES were made as follows: 0-2 Poor, 3-4 Fair, 5-6 Good, 7-8 Very Good and 9-10 Excellent.

Descriptive frequency statistics were also used to summarise the data, which were generated from document analysis using computerised SPSS, based on the above prescribed quantitative scales and scores stipulated under the PES. Summarised emerging frequency patterns supported the evaluation process in generating new insights into the behaviour of outcomes and assisted in creating meaningful interpretations of the results of the assessment based on emerging patterns.

General findings from document analysis without specific reference to specified compulsory associated evaluation standards based on the PES showed that the strengths of the first and fourth evaluation cases were found to be poor on the basis of the Stakeholder Identification (US1) criterion. On the basis of this criterion, the findings from the assessment of the strengths of selected evaluation cases from document analysis showed that a range of persons involved in, or affected by, the evaluations under consideration were not fully identified and their needs were not adequately addressed.

The findings from document analysis also revealed that persons who conducted the evaluation were not fully competent to perform the evaluation according to the Evaluator Credibility (US2) Criterion. Deficiencies in the level of trust and competency of evaluators may have undermined the credibility of evaluation outcomes and contributed significantly in exacerbating remarkable inconsistencies in the evaluation outcomes in the effectiveness of selected poverty-reduction interventions.

In terms of Defensible Information Sources (AS4) criteria, the findings revealed that the strengths of both the first and fourth evaluation cases were also found to be poor. Evident from the findings based on these criteria seemed to be that the sources of information, which were used in programme evaluation, were not described in enough detail to provide adequate information to support the evaluation process.

In addition, evidence from the study showed that the strengths of the second and third evaluation cases were poor on the Reliable Information (AS6) measure. On the basis of this measure, it was shown in the study that appropriate procedures of gathering information were not well selected and implemented to ensure that the information, which was obtained during the evaluation process, was sufficiently reliable for the intended use.

The study found that quantitative information was not appropriately or systematically analysed so that evaluation questions could be answered effectively. On the basis of the Analysis of Quantitative Information (AS8) measure, it was demonstrated in the study that the strengths of the first and third evaluation cases were also poor.

The PES stipulated under the Metaevaluation (AS12) that an evaluation itself should be formatively and summatively evaluated against these PES and other pertinent standards, so that its conduct is appropriately guided, and on completion, stakeholders could closely examine its strength and weaknesses. It was demonstrated in the thesis that the strengths of all the four selected evaluation cases were found to be poor on this criterion.

With reference to specified compulsory associated standards under the PES, however, it was revealed in the thesis that the strength of each of the four selected evaluation cases was not found to be poor, and therefore none of the selected evaluation cases could be rejected, based on compulsory associated evaluations standards specified under the PES.

The evidence from the study provided clear proof that the description and documentation of the programmes attracted the highest level of compliance, compared to: sources of defensible information; the process of systematically collecting, processing and reporting information; and Metaevaluation, which attracted lowest levels of compliance across the four selected evaluation cases. Evidence of new lessons that emerged from the study also showed that description and documentation of poverty reduction programmes were more important to the evaluation process than ensuring that the process of evaluation and the evaluation itself was formatively and summatively evaluated against pertinent standards; so that its conduct is appropriately guided, and, on completion, stakeholders could closely examine its strength and weaknesses.

The thesis also uncovered evidence showing that complying with practical procedures in the evaluation process, was more important than complying with the need to anticipate differing positions of various interest groups, in order to, obtain their cooperation, so that, possible attempts by any group to either curtail evaluation operations or misapply the results could be avoided. Evidence was also revealed from the study showing that, agreements in writing on obligations of formal parties to an evaluation took precedence over the obligation to openly and honestly deal with conflict of interest which could have, potentially compromised the evaluation processes and results.

Furthermore, the study illustrated that ensuring that evaluation reports clearly described the programme being evaluated, including its context, the purposes and findings of the evaluation, so that, essential information is provided and easily understood was of greater importance to the evaluation process, than, ensuring that persons who are involved in or affected by the evaluation needed to be identified, so that their needs could be addressed throughout the evaluation process.

In answering the research question [How effectively did evaluations, conducted under the auspices of the World Bank on poverty-reduction interventions funded through the PRSCs, meet the requirements of a good evaluation?], the study provided significant lessons, which demonstrated that there were evident variations in the extent to which each of the four selected evaluation cases complied with the prescribed requirements under the PES. Notwithstanding these variations in the extent to which each of the four selected evaluation cases complied with the specified compulsory associated standards of a good evaluation under the PES, it was demonstrated in this study that the strength and quality of each of the four selected evaluation cases was not found to be poor.

While the study revealed that the strength and quality of each of the four selected evaluation cases was not found to be poor, the evidence from the analysis revealed that the four cases performed better on bureaucratic criteria and worse on accountability and transparency procedures. The evidence of robust performance on bureaucratic criteria compared to the ineffective compliance with accountability and transparency procedures contradicted the raison d'etre for the intervention of the World Bank in Uganda, which was to promote accountability and transparency by trimming bloated bureaucracies.

In the study, I argued that the failure to promote accountability and transparency through trimming bloated bureaucracies in Uganda was attributed to both inappropriate development policy assumptions and prescriptions, which were pursued by the World Bank in support of Uganda’s reform process. The next section outlines the contribution to knowledge and suggests areas for further Metaevaluation research.

6.2 Contribution to Knowledge

The findings and conclusions from the study have generated practical implications on the methodological approach for conducting metaevaluation based on the PES. In particular, it is clarified in the findings and conclusions from the study that the quest to promote accountability and transparency within the context of evaluative enquiry is undergirded by participatory principles (Cousins and Earl 1992; Cousins and Whitmore 1998).

Participation is a form of collaboration, which involves various stakeholders, such as individuals, groups and communities that have an interest in the social development intervention (Cousins and Whitmore 1998, Owen and Rogers 1999). Participatory approaches for development have gained continuous rapid interest in many parts of the world. In particular, there has been heightened demand for the increased use of participatory approaches in designing and evaluating social development interventions in developing countries (Thompson, 1995; Cleaver, 1999; Declaration, 2005). Cleaver, (1999) for example, explains that participatory approaches broaden ownership, ensuring greater efficiency and effectiveness of development interventions.

Thompson (1995) also explains that participatory approaches tear down authoritarianism and inefficient bureaucracy, both of which are incapable of inspiring experimentation and self-critical learning or imaginative change. Furthermore, the Paris Declaration on Aid Effectiveness emphasises the importance of ownership, harmonisation, alignment, results and mutual accountability (Declaration, 2005).

Therefore, it is logical to argue that embedded in the theoretical postulations on accountability, transparency and social enquiry, as envisaged by Alkin (1969; 2012) and; Shadish, Cook and Leviton (1991), is an implicit demand for increased use of participatory approaches in designing and evaluating social development interventions, in order to improve the effectiveness of social development interventions sustainably.

In light of such heightened demand for increased use of participatory techniques, the existing specified compulsory associated standards of a good evaluation, based on the PES developed by the Joint Committee (1994) and grounded in the theory for good social programme evaluation, provide limited methodological scope and coverage towards assessing the quality of evaluations of social development interventions.

The limited methodological scope and coverage provided under the specified compulsory associated evaluation standards inescapably exacerbated these remarkable inconsistencies in evaluation outcomes of the effectiveness of development interventions. Therefore, it is plausible to argue within the context of participatory imperatives that social development interventions cannot be successfully implemented in the absence of adequate participation of a full range of stakeholders. The identification and engagement of a full range of stakeholders serves as a form of collaborative platform, which provides practical steps towards enhancing broad decision-making, thus improving accountability and transparency as well as solving the problem of bureaucratic inefficiency.

Contrary to heightened demand for increased use of participatory approaches in designing and evaluating social development interventions in the recent decade, the evidence from the analysis of the findings surprisingly revealed that a full range of stakeholders was not adequately identified and engaged in both designing, as well as in the evaluation process. The absence of the engagement of a full range of stakeholders undoubtedly contributes significantly to remarkable inconsistencies in the evaluation outcomes of the effectiveness of poverty reduction interventions. Confronted with this conundrum, the researcher makes one practical contribution to Metaevaluation methodology based on PES developed by the Joint Committee on PES in 1994 (Joint Committee 1994).

While stakeholder identification is not prescribed as one of the compulsory associated evaluation standards based on the PES developed in 1994 as above, to measure the quality of evaluations, it is empirically plausible to elevate Stakeholder Identification(US1) as an additional independent compulsory associated standard under Utility Standards based on the PES when conducting metaevaluations of evaluations of, not only poverty reduction interventions, but also other social development interventions.

The elevation of Stakeholder Identification (US1) as an additional independent compulsory associated standard under Utility Standards (US) based on the PES can play an important role of not only contributing towards a better resolution of plausible causes of inconsistencies in the evaluation outcomes of the effectiveness of social development interventions, but can also contribute to a better understanding of how, in practice, the identification and engagement of a full range of stakeholders enhances the promotion of the principles of good governance, such as accountability and transparency which are grounded in participatory theoretical imperatives. These principles, if applied, correctly can also insulate evaluation processes of social development interventions from other causes of inconsistencies, such as bureaucratic bottlenecks and conflict of interest, which also emerged from the analysis of the findings from this study.

In addition, the identification and engagement of a full range of stakeholders broadens the allocation of the exercise of power in the production of knowledge about the effectiveness of the project, programme, organisation and process. The identification and engagement of a full range of stakeholders also enhances broad ownership in the use of evaluation outcomes as well as the promotion of broad-based social change. It can also be added here that in light of this evidence, the centrality of the role played by stakeholders within the context of participatory evaluation is not only intended to enhance the effectiveness of development interventions, but can also if used properly constrain errant decision-makers from the exercise of power by holding them accountable - an exercise which cannot be underestimated or ignored.

The elevation of the important role, which Stakeholder Identification (US1) can play as an additional independent compulsory associated standard under Utility Standards (US) based on the PES, does not only contribute towards a better resolution of one of the plausible causes of remarkable inconsistencies in the evaluation outcomes of the effectiveness of social development interventions, but also contributes to a better understanding of how, in practice, the identification and engagement of a full range of stakeholders is a vital ingredient in enhancing the principles of good governance, such as accountability, transparency and participation, all of which not only insulate evaluation processes of social development interventions from remarkable inconsistencies, but may also ameliorate the adverse effects caused by both bureaucratic bottlenecks and conflict of interest.

The researcher hopes that this thesis has clarified and resolved one of the causes of the evident remarkable inconsistencies in evaluation outcomes of the effectiveness of poverty-reduction interventions in Uganda funded by the World Bank through its PRSCs. The researcher also hopes that the thesis has contributed to a better understanding of how metaevaluation helps in enhancing the ability to judge the strengths and weaknesses of a primary evaluation; validating a primary evaluation; adding credibility to the primary evaluation and improving the confidence of the users of the results of the primary evaluation so that interested parties can be able to judge the reliability and credibility of evaluations and the evidence used to make decisions.

6.3 Suggested Areas for Further Research

This thesis has particularly identified the need to engage a full range of stakeholders in designing and evaluating social development intervention as a panacea for resolving remarkable inconsistencies in evaluation outcomes of the effectiveness of social development interventions. However, the researcher proposes the following areas, which significantly emerged as problematic from the study for future research in metaevaluation and which the researcher could not continue undertaking due to logistical and conceptualisation limitations prescribed for purposes of this study.

6.3.1 Assessment of the Credibility of Evaluators: A metaevaluation of the competencies of persons conducting the evaluation and the identification of the nature of deficiencies in the level of professional trust in persons conducting the evaluation which may have potentially undermined the credibility of evaluation outcomes may contribute towards resolving possible causes in the effectiveness of evaluation outcomes and can also contribute to the improvement of the practice of the evaluation profession.

6.3.2 Assessment of Defensible Information Sources: An assessment of the reliability of information sources and the effectiveness with which such sources of information are described when conducting primary evaluations would be central in providing suitable information on the extent to which the evaluation process was supported with defensible sources of information.

6.3.3 Assessment of Reliable Information: An investigation into the use of appropriate procedures of gathering information during the evaluation process and the extent to which such information was sufficiently reliable for the intended user of the primary evaluation would contribute towards providing defensible grounds to inform decision-making processes.

6.3.4 Analysis of Quantitative Information: An examination of the degree of rigor with which appropriate and systematic quantitative information was analysed when conducting primary evaluations would contribute towards assisting in understanding the extent to which evaluation questions were answered effectively.

LIST OF REFERENCES

Ackoff, R. L. (1977). National development planning revisited. *Operations Research*, *25*(2), 207-218.

Agyemang, K. K. (2001). An overview of Structural Adjustment Programs in Africa. In K. Agyemang-Konadu (Ed.), *IMF and World Bank sponsored Structural Adjustment Programs in Africa*: *Ghana's Experiences* (pp.1983-1999). Farnham, England: Aldershot Publishing.

Alkin, M. C. (1969). *Evaluating the cost-effectiveness of instructional programmes*. Retrieved from https://eric.ed.gov/?id=ED031818.

Alkin, M. C. (Ed.). (2012). *Evaluation roots: A wider perspective of theorists' views and influences*. Thousand Oaks, CA: Sage Publications.

Amundsen, I. (2006). Political corruption and the role of donors (in Uganda). *Chr. Michelsen Institute Commissioned Report.* Kampala, Uganda: Royal Norwegian Embassy.

Appleton, S. (2001). *Poverty reduction during growth: The case of Uganda, 1992-2000*. Nottingham, England: Centre for Research in Economic Development and International Trade, School of Economics, University of Nottingham.

Apthorpe, R., & Gasper, D. (1982). Policy evaluation and meta-evaluation: The case of rural co-operatives. *World Development*, *10*(8), 651-668.

Ayittey, G. B. (1987). Economic atrophy in black Africa. *Cato Journal*, *7*(2), *195*—222.

Ayittey, G. B. (1992). *Africa betrayed.* New York, NY: St. Martin's Press.

Ayittey, G. B. (1994). *The Somali crisis: Time for an African solution*. Washington, DC: Cato Institute.

Ayittey, G. B. (1994). *The failure of development planning in Africa.* New York, NY: St. Martin's Press.

Ayittey, G. B. (1998). *Africa in chaos.* New York, NY: St. Martin's Press.

Ayittey, G. B. (2003).Why structural adjustment failed in Africa. *Contemporary economic issues in developing countries,* Wesport, CT: Praeger Publishing Group.

Ayittey, G.B. (2005). *Africa unchained. The blueprint for Africa's future*. New York, NY: Palgrave.

Ayittey, G. B. (2009). An African solution: Solving the crisis of failed states. *Harvard International Review*, *31*(3), 24.

Barkan, J. D. (2011). *Uganda: assessing risks to stability*. Washington, D.C: Center for Strategic and International Studies (CSIS).

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The Qualitative Report*, *13*(4), 544-559.

Bevir, M. (2008). *Key concepts in governance*. Thousand Oaks, CA: Sage Publications.

Beywl, W., & Speer, S. (2004). Data and literature-based reflections on Western European evaluation standards and practices. *New Directions for Evaluation*, *2004*(104), 43-54.

Borich, G. D. (1990). *Decision oriented evaluation*. Oxford, England: Pergamon Press.

Brett, E. A. (1996). *Rebuilding war-damaged communities in Uganda. In search of cool ground: War, flight, and homecoming in Northeast Africa.* Laurenceville, GA: Africa World Press.

Bigsten, A., & Kayizzi-Mugerwa, S. (1999). *Is Uganda an emerging economy? A report for the OECD project "emerging Africa"*. Göteborg, Sweden: Göteborg University.

Bissio, R. (2008). Paris declaration on aid effectiveness. Presentation at Human Rights Council, session of the Working Group on the Right to Development, High Task Force on the implementation of the Right to Development, Geneva, Switzerland: 7-15 January.

Bollen, K., Paxton, P., & Morishima, R. (2005). Assessing international evaluations: An example from USAID's democracy and governance program. *American Journal of Evaluation*, *26*(2), 189-203.

Chatterji, M. (2005). Applying The Joint Committee's 1994 Standards in international contexts: A case study of education evaluations in Bangladesh. *Teachers College Record*, *107*(10), 2372-2400.

Churchill, W. (2015). *My African journey (Vol. 8).* London, England: Richard Clay & Sons Limited.

Cleaver, F. (1999).Paradoxes of participation: Questioning participatory approaches to development. *Journal of International Development*, *11*(4), 597-612.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*(4), 397-418.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *1998*(80), 5-23.

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

Cooksy, L. J., & Caracelli, V. J. (2008). Meta-evaluation in practice: Selection and application of criteria. *Journal of Multidisciplinary Evaluation*, *6*(11), 1-15.

Cronbach, L. J. (2000). Course Improvement through Evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation Models: Viewpoints on Educational and Human Services* (2th ed., pp. 235-247). Hingham, MA: Kluwer Academic Publishers, Inc.

Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., ... & Weiner, S. S. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass Publishers.

De Long, J. B., & Summers, L. H. (1993). How strongly do developing economies benefit from equipment investment? *Journal of Monetary Economics*, *32*(3), 395-415.

Davidson, B. (1981). *The people's cause: A history of guerrillas in Africa*. Boston, MA: Addison-Wesley Publishing.

Demery, L. (1994). Structural adjustment: its origins, rationale and achievements. In A.G. Cornia & G.K. Helleiner (Eds.), *From Adjustment to Development in Africa* (pp. 25-48). London, England: Palgrave Macmillan Publishing.

Diamond, L. (2008). *The state of democracy in Africa*. *Democratization in Africa: What progress toward institutionalization?* Accra, Ghana: Ghana Center for Democratic Development.

Diamond, L. (2008). The rule of law versus the big man. *Journal of Democracy*, *19*(2), 138-149.

Doornbos, M. (1990). The African state in academic debate: Retrospect and Prospect. *The Journal of Modern African Studies*, *28*(2), 179-198.

Doornbos, M. (2005). *Transition and legitimacy in African states: The cases of Somalia and Uganda*. Retrieved from http://fride.org/download/WP17_SomaliaUga_ENG_dic05.pdf

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics*, *112*(4), 1203-1250.

Ffrench-Davis, R. (1988). *An outline of a neo-structuralist approach*. Santiago, Chile: Cepal Review.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines.* Boston, MA: Pearson Publishing.

Friedman, M., & Allen, G. (1970). The Counter-revolution in monetary theory: First Wincott memorial lecture delivered at the Senate House, University of London, London, England: *Institute of Economic Affairs*, *33(10, 1-28*.

Glass, G. V. (1969). *The growth of evaluation methodology*. Research paper no. 27, Boulder, Colo., University of Colorado, CO: Laboratory of Educational Research.

Glass, G.V., McGaw, B., & Smith, M.L. (1981*). Meta-analysis in social research*. Thousand Oaks, CA: Sage Publications.

Greene, J. C. (1992). A case study of evaluation auditing as meta-evaluation. *Evaluation and Program Planning*, *15*(1), 71-74.

Greene, J. C. (2001). Dialogue in evaluation: A relational perspective. *Evaluation, 7*(2), 181-187.

Grote, J. R., & Gbikpi, B. (2002). *Participatory governance*: *Political and societal implications.* Opladen, Germany: Leske and Budrich Publishers.

Guba, E. G. (1987). What have we learned about naturalistic evaluation? *Evaluation Practice*, *8*(1), 23-43.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Thousand Oaks, CA: Sage Publications.

Guba, E. G., & Lincoln, Y. S. (2001). *Guidelines and checklist for constructivist (aka fourth generation) evaluation.* Beverly Hills, CA: Sage Publications.

Hanssen, C. E., Lawrenz, F., & Dunet, D. O. (2008). Concurrent meta-evaluation: a critique. *American Journal of Evaluation*, *29*(4), 572-582.

Hedler, H., & Ribeiro, N. G. (2009). The contribution of meta-evaluation to program evaluation: Proposition of a model. *Journal of Multi-Disciplinary Evaluation*, *6*(12), 210-223.

Heyneman, S. P. (2003). The history and problems in the making of education policy at the World Bank 1960–2000. *International Journal of Educational Development*, *23*(3), 315-337.

Ilis, F., & Bahiigwa, G. (2003). Livelihoods and rural poverty reduction in Uganda. *World Development*, *31*(6), 997-1013.

International Covenant on Economic, Social and Cultural Rights, adopted 16 Dec. 1966, GA Res. 2200 (XXI), 21 UN GAOR Supp. (No. 16) at 49, UN Doc. A/6316 (1966), 993 UNTS 3.

Kapur, D., Lewis, J. P., & Webb, R. C. (2011). *The World Bank: its first half century* (Vol. 1).Washington, DC: Brookings Institution Press.

Kasfir, N. (2002). Dilemmas of popular support in guerrilla war: The national resistance army in Uganda–1981-86. *Journal of Modern African Studies*, *43*, 271-96.

Kaufmann, D. (2004). *Human rights and governance: The empirical challenge.* New York, NY: New York University School of Law.

Khiddu-Makubuya, E. (1994). Violence and conflict resolution in Uganda. *The Culture of Violence, 81*(1), 144-160.

Kifle, A. A. (2015). State-building and non-state conflicts in Africa 1. *Strategic Review for Southern Africa*, *37*(1), 99.

Knack, S., & Keefer, P. (1995). Institutions and economic performance: cross-country tests using alternative institutional measures. *Economics & Politics*, *7*(3), 207-227.

Kriger, Noma, J. (1992). *Zimbabwe guerilla war: Peasant voices.* Cambridge, England: Cambridge Press.

Leonard-Barton, D. (1990). A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science*, *1*(3), 248-266.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Newbury Park, CA: Sage Publications.

Mahle, N., Teferra, H., & Khachatryan, A. (2013). *Exchange rate liberalization in selected Sub-Saharan African countries successes, failures, and lessons.* Washington, DC: International Monetary Fund.

Mao, Z. (1975). *Selected works of Mao Zedong*. Sharpe, ME: Foreign Language Press.

Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 2(4), 81-712.

Mayne, J. (2001). Addressing attribution through contribution analysis: using performance measures sensibly. *The Canadian Journal of Program Ealuation*, *16*(1), 1.

Miller-Adams, M. (2002). *The World Bank: New agendas in a changing world*. New York, NY: Routledge Publishing

Mosley, P., Harrigan, J., & Toye, J. F. (1995). *Aid and power: the World Bank and policy-based lending* (Vol. 1). New York, NY: Psychology Press.

Mwenda, A. M., & Tangri, R. (2005). Patronage politics, donor reforms, and regime consolidation in Uganda. *African Affairs*, *104*(416), 449-467.

Mwenda, A. M. (2007). Personalizing power in Uganda. *Journal of Democracy*, *18*(3), 23-37.

Ndikumana, L., & Boyce, J. K. (2011). Capital flight from Sub-Saharan Africa: linkages with external borrowing and policy options. *International Review of Applied Economics*, *25*(2), 149-170.

Nkusu, M. M. (2004). *Financing Uganda's poverty reduction strategy: Is aid causing more pain than gain? (EPub)* (No. 4-170). Washington, D.C: International Monetary Fund.

North, D. C. (1982). *Structure & Change in Economic History*. New York, NY: WW Norton Publishing.

North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge, England: Cambridge University Press.

ODI. (2010). *Overseas Development Institute. Uganda: Case study for the MDG gap task force report.* London, England: London University Press.

Odom, S. L., & Fewell, R. R. (1983). Program evaluation in early childhood special education: A meta-evaluation. *Educational Evaluation and Policy Analysis*, *5*(4), 445-460.

OECD. (2004). *The organisation for Economic Co-operation and Development (OECD)*. Paris, France: Routledge Publishing.

Oestreich, J. E. (2004). The human rights responsibilities of the World Bank: A business paradigm. *Global Social Policy*, *4*(1), 55-76.

Orr, R. R., Cameron, S. J., & Day, D. M. (1991). Coping with stress in families with children who have mental retardation: An evaluation of the double ABCX model. *American Journal on Mental Retardation*, 95(4), 444-450.

Owen, J. M., & Rogers, P. (1999). *Programme evaluation: Forms and approaches*. Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (1978). *Utilization–focused evaluation*. Thousand Oaks, CA: Sage Publications

Patton, M. Q. (1997). Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice, 18*(2), 147-163.

Pierre, R. G. (1983). *Reanalyzing programme evaluation.* Basel, Switzerland: Jossey-Bass Publishers.

Rogers, P. J. (1992). The utilization-focused approach to evaluation: Does it ensure utilization? *Evaluation and Program Planning*, *18*(4), 321-332.

Rossi, P. H., Freeman, H.E & Lipsey, M.W. (1999). *Evaluation. A systematic approach.* Thousand Oaks, CA: Sage Publications.

Rueger, J. (2003). From reluctant champion to development ringmaster: Managing the expanding mission of the World Bank. *Indiana Journal of Global Legal Studies*, *10*(2), 201-225.

Russ-Eft, D & Preskill, H.S. (2001). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change*. New York, NY: Routledge Publishing.

Sala-i-Martin, X. (2006). The world distribution of income: falling poverty and… convergence, period. *The Quarterly Journal of Economics*, *121*(2), 351-397.

Salda, A. C. (1997). *Historical dictionary of the World Bank*. Lanham, MD: Scarecrow Press.

Sanders, J. R. (1994). *The program evaluation standards: how to assess evaluations of educational programs*. Thousand Oaks, CA: Sage Publications.

Sanders, J. R. (2007). *What the program evaluation standards say about social justice*. Michigan, MI: Evaluation Center, Western Michigan University.

Santiso, C. (2001). World Bank and good governance: good governance and aid effectiveness: the World Bank and conditionality. *Geo. Public Policy Rev.*, *7*(2), 1-137.

Schwandt, T. A., & Halpern, E. S. (1988). *Linking auditing and meta-evaluation: Enhancing quality in applied research*. Thousand Oaks, CA: Sage Publications.

Scott, A. (1970). *Insurgency*. Chapel Hill, NC: The University of North Carolina Press.

Scriven, M. (1966). *Primary philosophy*. Thousand Oaks, CA: Sage Publications.

Scriven, M. (1969). An introduction to meta-evaluation. Retrieved from *https://eric.ed.gov/?id=EJ003653.*

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage Publications.

Scriven, M. (2007). *Key evaluation checklist (KEC).* Thousand Oaks, CA: Sage Publications.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage Publications.

Stern, N., & Ferreira, F. (1997). *The World Bank as" intellectual actor."* Washington D.C.: Brookings Institution.

Stake, R.E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage Publications.

Stake, R. E. (2006). *Multiple Case study Analysis*. New York, NY: Guilford Press.

Stige, B., Malterud, K., & Midtgarden, T. (2009). Toward an agenda for evaluation of qualitative research. *Qualitative Health Research*, *19*(10), 1504-1516.

Straus, A, and Corbin, J. (1990). *Basis of Qualitative Research: Grounded Theory of Qualitative Research*. Thousand Oaks, CA: Sage Publications.

Stufflebeam, D. L. (1971). *The relevance of the CIPP evaluation model for educational accountability*. Abingdon, England: Routledge Publishing.

Stufflebeam, D. L. (1974). *Meta-evaluation*. Kalamazoo, MI: Western Michigan University Evaluation Center.

Stufflebeam, D. L. (1978). Meta evaluation: an overview. *Evaluation and the Health Professions*, *1*(1), 17-43.

Stufflebeam, D. L. (1981). Meta-evaluation: Concepts, standards, and uses. *Educational Evaluation Methodology: The State of the Art*, 6(1), 146-163.

Stufflebeam, D. L. (1999). *Program evaluations meta-evaluation checklist*. Dordrecht, Netherlands: Springer Publishing.

Stufflebeam, D. L. (2000). *Foundational models for 21st century program evaluation*. Dordrecht, Netherlands: Springer Publishing.

Stufflebeam, D. L. (2001). The meta-evaluation imperative. *American Journal of Evaluation* 22(2), 183 - 209.

Stufflebeam, D.L Foley, W.J., Gephardt, W.J., Guba, E.G., Hammond, R.L., Merriman, H.D., & Provus, M.M. (1971). *Educational Evaluation and Decision Making*. San Francisco, CA: Peacock Publishers.

Stufflebeam, D.L & Shinkfield, A. (2007*). Evaluation theory, models and applications.* San Francisco: CA: John Wiley. Thousand Oaks, CA: Sage Publications.

Stufflebeam, D. L., & Webster, W. J. (1983). An analysis of alternative approaches to evaluation. In D.L. Stufflebeam (Ed.) *Evaluation models* (pp. 23-43). Dordrecht, Netherlands: Springer Publishing.

Tangri, R. (2005). *Politics and presidential term limits in Uganda: Legacies or power.* Cape Town, South Africa: Human Sciences Research Council Press.

Tangri, R., & Mwenda, A. M. (2010). President Museveni and the politics of presidential tenure in Uganda. *Journal of Contemporary African Studies*, *28*(1), 31-49.

The Joint Committee on Standards for educational Evaluation. (1994). *The programme evaluation standards* (2nd ed). Thousand Oaks, CA: Sage Publications.

The Ten Point Programme. (1985). In Yoweri Museveni (Ed.) *Selected articles on Uganda resistance war.* Kampala, Uganda: NRM Publications.

The Paris Declaration on Aid Effectiveness. (2005). Ownership, harmonisation, alignment, results and mutual accountability. In *High level forum on joint progress toward enhanced aid effectiveness: Harmonization, alignment and results. 2nd High level forum on aid effectiveness. Paris, France.*

Thompson, J. (1995). Participatory approaches in government bureaucracies: Facilitating the process of institutional change. *World Development*, *23*(9), 1521-1554.

Trebesch, C, Papaioannou, M & Uaibir, D. (2012). *Sovereign debt restructurings 1950-2010: literature survey, data and stylized facts.* Washington, DC: International Monetary Fund.

Toye, J. (1991). Is there a new political economy of development? In C. Colclough & J. Manor (Eds), *States or Markets? Neo-liberalism and the Development Policy Debate* (pp. 36-47). London, England: Clarendon Press.

Toye, J. (1993). *Structural adjustment: context, assumptions, origin and diversity*. Paper presented at the DGIS. /ISSAS Conference on Structural Adjustment and Beyond. The Hague, Netherlands.

Tyler, R. W. (1942). General statement on evaluation. *The Journal of Educational Research*, *35*(7), 492-501.

Herrick, V. E., & Tyler, R. W. (Eds.). (1950). *Towards improved curriculum theory: papers presented at the conference*. Chicago, IL: University of Chicago Press.

Uganda Bureau of Statistics. (2004). *Where are the poor? Mapping patterns of well-being in Uganda: 1992 and 1999*. Kampala, Uganda: NRM Publications.

United Nations Development Programme. (1980). C*ountry and intercountry programming and project*s. New York, NY: United Nations Publications.

United Nations. (2002).*World Population Prospects*. Economic and Social Affairs, Volume 3, New York, NY: United Nations Publications.

United Nations. (2008). *The Millennium Development Goals Report 2008.* New York, NY: United Nations Publications.

Uusikylä, P., & Virtanen, P. (2000). Meta-evaluation as a tool for learning: A case study of the European structural fund evaluations in Finland. *Evaluation*, *6*(1), 50-65.

Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, *8*(1), 60-80.

Weaver, J. H. (1989). Competing Paradigms of Development. *Social Education*, *53*(4), 209-13.

Weiss, C. H. (1972). Utilization of evaluation: Toward comparative study. *Evaluating Action Programs: Readings in Social Action and Education*, 6(3), 318-326.

Weiss, C.H. (1983). Ideology, interest, and information: The basis of policy decisions. In D. Callahan & B. Jennings (Eds.), *Ethics, the social sciences, and policy analysis* (pp.213-245). New York, NY: Plenum Publishing.

Weiss, C.H. (1998). *Evaluation*. Upper Saddle, NJ: Prentice Hall Publishing.

Whitaker, J. (1988). *How can Africa survive*? New York, NY: Harper & Row Publishing.

Wholey, J. S. (2001). Managing for results: Roles for evaluators in a new management era. *American Journal of Evaluation*, *22*(3), 343-347.

Wholey, J.S, Hatry, H.P & Newcomer, K.E (Eds.) (2004). *Handbook of practical programme evaluation*. San Francisco, CA: Jossey-Bass Publishers.

Wingate, L. A., Coryn, C. L. S., Gullickson, A. R., & Cooksy, L. J. (2010). The Program Evaluation Standards applied for meta-evaluation purposes: Investigating interrater agreement and its implications for use. *Manuscript submitted for publication*.

Wittrock, M. C., & Wiley, D. E. (1970). *The evaluation of instruction: Issues and problems*. New York, NY: Holt, Rinehart and Winston Publishing.

Williams, Z. T. (2002). *Africa in crisis*. London, England: Pluto Publishing.

Williamson, J. (1990). *The progress of policy reform in Latin America*. Washington, DC: Institute for International Affairs.

Williamson, J. (2000). What should the World Bank think about the Washington Consensus? *The World Bank Research Observer*, *15*(2), 251-264.

Wingate, L.A. (2010). Meta-evaluation: Purpose, prescription, and practice. In E. Bader, P. Peterson, & B. McGraw (Eds.), *International Encyclopedia of Education* (3rd ed.) (pp. 765-774). San Diego, CA: Elsevier Publishing.

World Bank. (1989). *Sub-Saharan Africa: From crisis to sustainable growth*. Washington, DC: World Bank Group.

World Bank. (1997).*Adjustment in Africa: Reforms, Results and the Road Ahead*. New York, NY: Oxford University Press.

World Bank. (2006). *Uganda: Poverty and vulnerability assessment report*. Washington, DC: World Bank Group.

World Bank. (2009a). *Uganda: Project performance assessment report*. Washington, DC: World Bank Group.

World Bank. (2009b*). Uganda: First Poverty Reduction Support Credits (Credit Number 3510): Project performance assessment report.* Washington, DC: World Bank Group.

World Bank. (2009c). *Uganda: Second poverty reduction support credits (Credit Number 3696): Project performance assessment report*. Washington, DC: World Bank Group.

World Bank. (2009d). *Uganda: Third poverty reduction support credits (Grant Number. H0630): Project performance assessment report*. Washington, DC: World Bank Group.

World Bank. (2009e). *Uganda: Fourth poverty reduction support credits (Grant Number. H1220-UG): Project performance assessment report*. Washington, DC: World Bank Group.

World Bank. (2010). *African development indicators 2008*. Washington, DC: World Bank Publications.

World Bank. (2011a). *African development indicators 2010*. Washington, DC: World Bank Publications.

World Bank. (2011b). *Africa action plan: IEG evaluation, World Bank.* Washington, DC: World Bank Group.

Yilmaz, K. (2008). Constructivism: Its theoretical underpinnings, variations, and implications for classroom instruction. *Educational horizons*, *86*(3), 161-172.

Yin, R. K. (1989). Research design issues in using the case study method to study management information systems. *The information systems research challenge: Qualitative research methods*, *1*(6), 1-6.

Yin, R. K. (1993). *Applications of case study research. Applied social research methods series*. Thousand Oaks, CA: Sage Publications.

Yin, R. K. (2009). *Case study research: Design and methods*, (4th ed.).Thousand Oaks, CA: Sage Publications.

APPENDIX A

Utility Standards

Table A1: Utility Standards (US)

To Meet The Requirements For Utility, Programme Evaluations SHOULD:	
US 1: Stakeholder Identification	
	Clearly identify the evaluation client.
	Engage leadership figures to identify other stakeholders.
	Consult stakeholders to identify their information needs.
	Use stakeholders to identify other stakeholders.
	With the client, rank stakeholders for relative importance.
	Arrange to involve stakeholders throughout the evaluation, consistent with the formal evaluation agreement.
	Keep the evaluation open to serve newly-identified stakeholders.
	Address stakeholders' evaluation needs.
	Serve an appropriate range of individual stakeholders.
	Serve an appropriate range of stakeholder organisations.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
US2: Evaluator Credibility	
	Engage competent evaluators.
	Engage evaluators whom the stakeholders trust.
	Engage evaluators that can address stakeholders' concerns.
	Engage evaluators that are appropriately responsive to issues of gender, socio-economic status, race, and language and cultural differences.
	Assure that the evaluation plan responds to key stakeholders' concerns.
	Help stakeholders to understand the evaluation plan.
	Give stakeholders information on the evaluation plan's technical quality and practicality.
	Attend appropriately to stakeholders' criticisms and suggestions.
	Stay abreast of social and political forces.
	Keep interested parties informed about the evaluation's progress.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table A1: (Continued)

US3: Information Scope and Selection	
	Understand the client's most important evaluation requirements.
	Interview stakeholders to determine their different perspectives.
	Assure that evaluator and client negotiate pertinent audiences, questions and required information.
	Assign priority to the most important stakeholders.
	Assign priority to the most important questions.
	Allow flexibility for adding questions during the evaluation.
	Obtain sufficient information to address the stakeholders' most important evaluation questions.
	Obtain sufficient information to assess the programme's merit.
	Obtain sufficient information to assess the programme's worth.
	Allocate the evaluation effort in accordance with the priorities assigned to the needed information.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
US4: Values Identification	
	Consider alternative sources of values for interpreting evaluation findings.
	Provide a clear, defensible basis for value judgments.
	Determine the appropriate party(ies) to make the valuational interpretations.
	Identify pertinent societal needs
	Identify pertinent customer needs.
	Reference pertinent laws.
	Reference, as appropriate, the relevant institutional mission.
	Reference the programme's goals.
	Take into account the stakeholders' values.
	As appropriate, present alternative interpretations based on conflicting but credible value bases.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table A1: (Continued)

US5: Report Clarity	
	Clearly report the essential information.
	Issue brief, simple and direct reports.
	Focus reports on contracted questions.
	Describe the programme and its context.
	Describe the evaluation's purposes, procedures and findings.
	Support conclusions and recommendations.
	Avoid reporting technical jargon.
	Report in the language(s) of stakeholders.
	Provide an executive summary.
	Provide a technical report.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
US6: Report Timeliness and Dissemination	
	Make timely interim reports to intended users.
	Deliver the final report when it is needed.
	Have timely exchanges with the programme's policy board.
	Have timely exchanges with the programme's staff.
	Have timely exchanges with the programme's customers.
	Have timely exchanges with the public media.
	Have timely exchanges with the full range of right-to-know audiences.
	Employ effective media for reaching and informing the different audiences.
	Keep the presentations appropriately brief.
	Use examples to help audiences relate the findings to practical situations.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table A1: (Continued)

US7: Evaluation Impact	
	Maintain contact with audience.
	Involve stakeholders throughout the evaluation.
	Encourage and support stakeholders’ use of the findings.
	Show stakeholders how they might use the findings in their work.
	Forecast and address potential uses of findings.
	Provide interim reports.
	Make sure that reports are open, frank and concrete.
	Supplement written reports with ongoing oral communication.
	Conduct feedback workshops to go over, and apply, findings.
	Make arrangements to provide follow-up assistance in interpreting and applying the findings.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
Scoring the Evaluation for UTILITY Add the following: Number of Excellent ratings (0-7) …...x 4 = ….. Number of Very Good (0-7) …...x 3 =…... Number of Good (0-7)…… x 2 =….. Number of Fair (0-7)…… x 1 =… Total Score ………………………………	Strength of the evaluation’s provisions for UTILITY: 26 (93%) to 28: Excellent 19 (68%) to 25: Very Good 14 (50%) to 18: Good 7 (25%) to 13: Fair 0 (0%) to 5: Poor (Total score) ÷28 = __ x 100 = ___

Sources: Sanders, J. R. (1994) & Stufflebeam, D. L. (1999).

APPENDIX B

Feasibility Standards

Table B1: Feasibility Standards (FS) Checklist.

To Meet The Requirements For FEASIBILITY, Programme Evaluations SHOULD:	
FS1 Practical Procedures	
	Tailor methods and instruments to information requirements.
	Minimise disruption.
	Minimise the data burden.
	Appoint competent staff.
	Train staff.
	Choose procedures that the staff are qualified to carry out.
	Choose procedures in light of known constraints.
	Make a realistic schedule.
	Engage locals to help conduct the evaluation.
	As appropriate, make evaluation procedures a part of routine events.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
FS2 Political Viability	
	Anticipate different positions of different interest groups.
	Avert or counteract attempts to bias or misapply the findings.
	Foster cooperation.
	Involve stakeholders throughout the evaluation.
	Agree on editorial and dissemination authority.
	Issue interim reports.
	Report divergent views.
	Report to right-to-know audiences.
	Employ a firm public contract.
	Terminate any corrupted evaluation.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table B1: (Continued)

FS3 Cost Effectiveness	
	Be efficient.
	Make use of in-kind services.
	Produce information worth the investment.
	Inform decisions.
	Foster programme improvement.
	Provide accountability information.
	Generate new insights.
	Help spread effective practices.
	Minimise disruptions.
	Minimise time demands on programme personnel.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
Scoring the Evaluation for FEASIBILITY Add the following: Number of Excellent ratings (0-3)…….x 4 = ….. Number of Very Good (0-3) …...x 3 =…... Number of Good (0-3)…… x 2 =….. Number of Fair (0-3)…… x 1 =…… Total Score ………………………………	Strength of the evaluation's provisions for FEASIBILITY: 11 (93%) to 12: Excellent 8 (68%) to 10: Very Good 6 (50%) to 7: Good 3 (25%) to 5: Fair 0 (0%) to 2: Poor (Total score) ÷12 =__ x 100 = __

Sources: Sanders, J. R. (1994) & Stufflebeam, D. L. (1999).

APPENDIX C

Propriety Standards

Table C1: Propriety Standards (PS) Checklist.

To Meet The Requirements For PROPRIETY, Programme Evaluations SHOULD:	
PS1 Service Orientation	
	Assess needs of the programme's customers.
	Assess programme outcomes against targeted customers' assessed needs.
	Help to assure that the full range of rightful programme beneficiaries is served.
	Promote excellent service.
	Make the evaluation's service orientation clear to stakeholders.
	Identify programme strengths on which to build.
	Identify programme weaknesses to correct.
	Give interim feedback for programme improvement.
	Expose harmful practices.
	Inform all right-to-know audiences of the programme's positive and negative outcomes.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
PS2 Formal Agreements - Reach Advance written Agreements on:	
	Evaluation purpose and questions.
	Audiences.
	Evaluation reports.
	Editing.
	Release of reports.
	Evaluation procedures and schedule.
	Confidentiality/anonymity of data.
	Evaluation staff.
	Metaevaluation.
	Evaluation resources.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
PS3 Rights of Human Subjects	
	Make clear to stakeholders that the evaluation will respect and protect the rights of human subjects.
	Clarify intended uses of the evaluation.
	Keep stakeholders informed.
	Follow due process.
	Uphold civil rights.
	Understand participant values.
	Respect diversity.
	Follow protocol.
	Honour confidentiality/anonymity agreements.
	Do no harm.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table C1: (Continued)

PS4 Human Interactions	
	Consistently relate to all stakeholders in a professional manner.
	Maintain effective communication with stakeholders.
	Follow the institution's protocol.
	Minimise disruption.
	Honour participants' privacy rights.
	Honour time commitments.
	Be alert to, and address, participants' concerns about the evaluation.
	Be sensitive to participants 'diversity of values and cultural differences.
	Be even-handed in addressing different stakeholders.
	Do not ignore or help to cover up any participant's incompetence, unethical behavior, fraud, waste or abuse.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
PS5 Complete and Fair Assessment	
	Assess and report the programme's strengths.
	Assess and report the programme's weaknesses.
	Report on intended outcomes.
	Report on unintended outcomes.
	Give a thorough account of the evaluation's process.
	As appropriate, show how the programme's strengths could be used to overcome its weaknesses.
	Have the draft report reviewed.
	Appropriately address criticisms of the draft report.
	Acknowledge the final report's limitations
	Estimate and report the effects of the evaluation's limitations on the overall judgment of the programme.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
PS6 Disclosure of Findings	
	Define the right-to-know audiences.
	Establish a contractual basis for complying with right-to-know requirements.
	Inform the audiences of the evaluation's purposes and projected reports.
	Report all findings in writing.
	Report relevant points of view of both supporters and critics of the programme.
	Report balanced, informed conclusions and recommendations.
	Show the basis for the conclusions and recommendations.
	Disclose the evaluation's limitations.
	In reporting, adhere strictly to a code of directness, openness, and completeness.
	Assure that reports reach their audiences.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table C1: (Continued)

PS7 Conflict of Interest	
	Identify potential conflicts of interest early in the evaluation.
	Provide written, contractual safeguards against identified conflicts of interest.
	Engage multiple evaluators.
	Maintain evaluation records for independent review.
	As appropriate, engage independent parties to assess the evaluation for its susceptibility or corruption by conflicts of interest.
	When appropriate, release evaluation procedures, data and reports for public review.
	Contract with the funding authority rather than the funded programme.
	Have internal evaluators report directly to the chief executive officer.
	Report equitably to all right-to-know audiences.
	Engage uniquely-qualified persons to participate in the evaluation, even if they have a potential conflict of interest; but take steps to counteract the conflict.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
PS8 Fiscal Responsibility	
	Specify and budget for expense items in advance.
	Keep the budget sufficiently flexible to permit appropriate reallocations to strengthen the evaluation.
	Obtain appropriate approval for necessary budgetary modifications.
	Assign responsibility for managing the evaluation finances.
	Maintain accurate records of sources of funding and expenditure.
	Maintain adequate personnel records concerning job allocations and time spent on the job.
	Employ comparison shopping for evaluation materials.
	Employ comparison contract bidding.
	Be frugal in expending evaluation resources.
	As appropriate, include an expenditure summary as part of the public evaluation report.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
Scoring the Evaluation for PROPRIETY **Add the following:** Number of Excellent ratings (0-8)…….x 4 = ….. Number of Very Good (0-8) …...x 3 =…... Number of Good (0-8)…… x 2 =….. Number of Fair (0-8)…… x 1 =… **Total Score …………………………**	**Strength of the evaluation's provisions for PROPRIETY:** 30 (93%) to 32: **Excellent** 22 (68%) to 29: **Very Good** 16 (50%) to 21: **Good** 8 (25%) to 15: **Fair** 0 (0%) to 7: **Poor** **(Total score) ÷32 = __ x 100 = __**

Sources: Sanders, J. R. (1994) & Stufflebeam, D. L. (1999).

APPENDIX D

Accuracy Standards

Table D1: Accuracy Standards (AS) Checklist.

To Meet The Requirements For ACCURACY, Programme Evaluations SHOULD:	
AS1 Programme Documentation	
	Collect descriptions of the intended programme from various written sources.
	Collect descriptions of the intended programme from the client and various stakeholders.
	Describe how the programme was intended to function.
	Maintain records from various sources of how the programme operated.
	As feasible, engage independent observers to describe the programme's actual operations.
	Describe how the programme actually functioned.
	Analyse discrepancies between the various descriptions of how the programme was intended to function.
	Analyse discrepancies between how the programme was intended to operate and how it actually operated.
	Ask the client and various stakeholders to assess the accuracy of recorded descriptions of both the intended and the actual programme.
	Produce a technical report that documents the programme's operations.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS2 Context Analysis	
	Use multiple sources of information to describe the programme's context.
	Describe the context's technical, social, political, organisational and economic features.
	Maintain a log of unusual circumstances.
	Record instances in which individuals or groups, intentionally or otherwise, interfered with the programme.
	Record instances in which individuals or groups, intentionally or otherwise, gave special assistance to the programme.
	Analyse how the programme's context is similar to, or different from, contexts where the programme might be adopted.
	Report those contextual influences that appeared to influence the programme significantly and that might be of interest to potential adopters.
	Estimate effects of context on programme outcomes.
	Identify and describe any critical competitors to this programme that functioned at the same time and in the programme's environment.
	Describe how people in the programme's general area perceived the programme's existence, importance and quality.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Goo☐ 3-4☐ir 0-2 Poor	

Table D1: (Continued).

AS3 Described Purposes and Procedures	
	At the evaluation's outset, record the client's purposes for the evaluation.
	Monitor and describe stakeholders' intended uses of evaluation findings.
	Monitor and describe how the evaluation's purposes stay the same or change over time.
	Identify and assess points of agreement and disagreement among stakeholders regarding the evaluation's purposes.
	As appropriate, update evaluation procedures to accommodate changes in the evaluation's purposes.
	Record the actual evaluation procedures, as implemented.
	When interpreting findings, take into account the different stakeholders' intended uses of the evaluation.
	When interpreting findings, take into account the extent to which the intended procedures were effectively executed.
	Describe the evaluation's purposes and procedures in the summary and full-length evaluation reports.
	As feasible, engage independent evaluators to monitor and evaluate the evaluation's purposes.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS4 Defensible Information Sources	
	Obtain information from a variety of sources.
	Use pertinent, previously-collected information once validated.
	As appropriate, employ a variety of data collection methods.
	Document and report information sources.
	Document, justify and report the criteria and methods used to select information sources.
	For each source, define the population.
	For each population, as appropriate, define any employed sample.
	Document, justify and report the means used to obtain information from each source.
	Include data collection instruments in a technical appendix to the evaluation report.
	Document and report any biasing features in the obtained information.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table D1: (Continued).

AS5 Valid Information	
	Focus the evaluation on key questions.
	As appropriate, employ multiple measures to address each question.
	Provide a detailed description of the constructs and behaviours about which information will be acquired.
	Assess and report the type of information each employed procedure acquires.
	Train and calibrate the data collectors.
	Document and report the data collection conditions and process.
	Document how information from each procedure was scored, analysed and interpreted
	Report and justify inferences singly and in combination.
	Assess and report the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions.
	Establish meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table D1: (Continued).

AS6 Reliable Information	
	Identify and justify the type(s) and extent of reliability claimed.
	For each employed data collection device, specify the unit of analysis.
	As feasible, choose measuring devices that in the past have shown acceptable levels of reliability for their intended uses.
	In reporting reliability of an instrument, assess and report the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions and the evaluator's bias.
	Check and report the consistency of scoring, categorisation and coding.
	Train and calibrate scorers and analysts to produce consistent results.
	Pilot test new instruments in order to identify and control sources of error.
	As appropriate, engage and check the consistency between multiple observers.
	Acknowledge reliability problems in the final report.
	Estimate and report the effects of unreliability in the data on the overall judgment of the programme.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS7 Systematic Information	
	Establish protocols for quality control of the evaluation information.
	Train the evaluation staff to adhere to the data protocols.
	Systematically check the accuracy of scoring and coding.
	When feasible, use multiple evaluators and check the consistency of their work.
	Verify data entry.
	Proofread and verify data tables generated from computer output or other means.
	Systematise and control storage of the evaluation information.
	Define those who will have access to the evaluation information.
	Strictly control access to the evaluation information according to established protocols.
	Have data providers verify the data they submitted.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fa☐ 0-2 Poor	

Table D1: (Continued).

AS8 Analysis of Quantitative Information	
	Begin by conducting preliminary exploratory analyses to assure the data's correctness and to gain a greater understanding of the data.
	Choose procedures appropriate for the evaluation questions and nature of the data.
	For each procedure specify how its key assumptions are being met.
	Report limitations of each analytic procedure, including failure to meet assumptions.
	Employ multiple analytic procedures to check on consistency and replicability of findings.
	Examine variability as well as central tendencies.
	Identify and examine outliers and verify their correctness.
	Identify and analyse statistical interactions.
	Assess statistical significance and practical significance.
	Use visual displays to clarify the presentation and interpretation of statistical results.
☐ 9-10 Excellent ☐7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐-2 Poor	

Table D1: (Continued).

AS9 Analysis of Qualitative Information	
	Focus on key questions.
	Define the boundaries of information to be used.
	Obtain information keyed to the important evaluation questions.
	Verify the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders.
	Choose analytic procedures and methods of summarisation that are appropriate to the evaluation questions and employed qualitative information.
	Derive a set of categories that is sufficient to document, illuminate and respond to the evaluation questions.
	Test the derived categories for reliability and validity.
	Classify the obtained information into the validated analysis categories.
	Derive conclusions and recommendations and demonstrate their meaningfulness.
	Report limitations of the referenced information, analyses and inferences.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS10 Justified Conclusions	
	Focus conclusions directly on the evaluation questions.
	Accurately reflect the evaluation procedures and findings.
	Limit conclusions to the applicable time periods, contexts, purposes and activities.
	Cite the information that supports each conclusion.
	Identify and report the programme's side effects.
	Report plausible alternative explanations of the findings.
	Explain why rival explanations were rejected.
	Warn against making common misinterpretations.
	Obtain and address the results of a pre-release review of the draft evaluation report.
	Report the evaluation's limitations.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS11 Impartial Reporting	
	Engage the client to determine steps to ensure fair, impartial reports.
	Establish appropriate editorial authority.
	Determine right-to-know audiences.
	Establish and follow appropriate plans for releasing findings to all right-to-know audiences.
	Safeguard reports from deliberate or inadvertent distortions.
	Report perspectives of all stakeholder groups.
	Report alternative plausible conclusions.
	Obtain outside audits of reports.
	Describe steps taken to control bias.
	Participate in public presentations of the findings to help to guard against, and to correct distortions by others.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	

Table D1: (Continued).

AS6 Reliable Information	
	Identify and justify the type(s) and extent of reliability claimed.
	For each employed data collection device, specify the unit of analysis.
	As feasible, choose measuring devices that in the past have shown acceptable levels of reliability for their intended uses.
	In reporting reliability of an instrument, assess and report the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions, and the evaluator's bias.
	Check and report the consistency of scoring, categorisation and coding.
	Train and calibrate scorers and analysts to produce consistent results.
	Pilot test new instruments in order to identify and control sources of error.
	As appropriate, engage and check the consistency between multiple observers.
	Acknowledge reliability problems in the final report.
	Estimate and report the effects of unreliability in the data on the overall judgment of the programme.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS7 Systematic Information	
	Establish protocols for quality control of the evaluation information.
	Train the evaluation staff to adhere to the data protocols.
	Systematically check the accuracy of scoring and coding.
	When feasible, use multiple evaluators and check the consistency of their work.
	Verify data entry.
	Proofread and verify data tables generated from computer output or other means.
	Systematise and control storage of the evaluation information.
	Define those who will have access to the evaluation information.
	Strictly control access to the evaluation information according to established protocols.
	Have data providers verify the data they submitted.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐5-6 Good ☐3-4 Fair ☐0-2 Poor	

Table D1: (Continued).

AS8 Analysis of Qualitative Information	
	Begin by conducting preliminary exploratory analyses to assure the data's correctness and to gain a greater understanding of the data.
	Choose procedures appropriate for the evaluation questions and nature of the data.
	For each procedure specify how its key assumptions are being met.
	Report limitations of each analytic procedure, including failure to meet assumptions.
	Employ multiple analytic procedures to check on consistency and replicability of findings.
	Examine variability as well as central tendencies.
	Identify and examine outliers and verify their correctness.
	Identify and analyse statistical interactions.
	Assess statistical significance and practical significance.
	Use visual displays to clarify the presentation and interpretation of statistical results.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
AS12 Metaevaluation	
	Designate or define the standards to be used in judging the evaluation.
	Assign someone responsibility for documenting and assessing the evaluation process and products.
	Employ both formative and summative Metaevaluation.
	Budget appropriately and sufficiently for conducting the Metaevaluation.
	Record the full range of information needed to judge the evaluation against the stipulated standards.
	As feasible, contract for an independent Metaevaluation.
	Determine and record those audiences that will receive the Metaevaluation report.
	Evaluate the instrumentation, data collection, data handling, coding and analysis against the relevant standards.
	Evaluate the evaluation's involvement of, and communication of, findings to stakeholders against the relevant standards.
	Maintain a record of all Metaevaluation steps, information and analyses.
☐ 9-10 Excellent ☐ 7-8 Very Good ☐ 5-6 Good ☐ 3-4 Fair ☐ 0-2 Poor	
Scoring the Evaluation for ACCURACY Add the following: Number of Excellent ratings (0-12)…….x 4 = ….. Number of Very Good (0-12) …...x 3 =…... Number of Good (0-12)…… x 2 =….. Number of Fair (0-12)…… x 1 =…… Total Score ………………………	Strength of the evaluation's provisions for ACCURACY: 45 (93%) to 48: Excellent 33 (68%) to 44: Very Good 24 (50%) to 32: Good 12 (25%) to 23: Fair 0 (0%) to 11: Poor (Total score) ÷ 48 = __ x 100 = ___

Sources: Sanders, J. R. (1994) & Stufflebeam, D. L. (1999).

APPENDIX E

Detailed Description of Programme Evaluation Standards

1.0 Utility Standards (US1-US7): These standards are meant to ensure that the information generated from evaluations serve the needs of intended users (Stufflebeam (2001, 2005), Rogers (1992), Weiss (1983), Patton (1978, 1986, and 1997) by paying attention to the following associated standards: stakeholder identification, evaluator credibility, information scope and selection, values identification, report clarity, report timeliness and dissemination, and report impact.

1.1 Stakeholder Identification (US1): provides that: "persons involved in or affected by the evaluation should be identified, so that, their needs can be addressed" (Sanders, 1994, p.23). Stufflebeam (2005) provides the following criteria for meeting this standard namely:

1 Clearly identify the evaluation client.
2 Engage leadership figures to identify other stakeholders.
3 Consult potential stakeholders to identify their information needs.
4 Use stakeholders to identify other stakeholders.
5 With the client, rank stakeholders for relative importance.
6 Arrange to involve stakeholders throughout the evaluation.
7 Keep the evaluation open to serve newly-identified stakeholders.
8 Address stakeholders' evaluation needs.
9 Serve an appropriate range of individual stakeholders.
10 Serve an appropriate range of stakeholder organisations.

1.2 Evaluator Credibility (US2): "The person conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that, the evaluation findings achieve maximum credibility and acceptance" (Sanders, 1994, p.23). Stufflebeam (2005) outlines the following criteria for meeting this standard:

1. Engage competent evaluators.
2. Engage evaluators trusted the stakeholders.
3. Engage evaluators that can address stakeholders' concerns.
4. Engage evaluators that are appropriately responsive to issues of gender, socioeconomic status, race, and language and cultural differences.
5. Ensure that the evaluation plan responds to key stakeholders' concerns.
6. Help stakeholders to understand the evaluation plan.
7. Give stakeholders information on the evaluation plan's technical quality and practicality.
8. Attend appropriately to stakeholders' criticisms and suggestions.
9. Stay abreast of social and political forces.
10. Keep interested parties informed about the evaluation's progress.

1.3 Information Scope and Selection (US3): "Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of the clients and other specified stakeholders" (Sanders, 1994, p.23). Stufflebeam (2005) specified the following criteria for meeting this standard:

1. Understand the client's most important evaluation requirements.
2. Interview stakeholders to determine their different perspectives.
3. Ensure that evaluator and client negotiate pertinent audiences, questions and required information.
4. Assign priority to the most important stakeholders.
5. Assign priority to the most important questions.
6. Allow flexibility for adding questions during the evaluation.
7. Obtain sufficient information to address the stakeholders' most important evaluation questions.
8. Obtain sufficient information to assess the programme's merit.
9. Obtain sufficient information to assess the programme's worth.
10. Allocate the evaluation effort in accordance with the priorities assigned to the required information.

1.4 Values Identification (US4): "The perspectives, procedures and rationale used to interpret the findings should be carefully described, so that, the bases for value judgments are clear" (Sanders, 1994, p.23). The following criteria are described by Stufflebeam (2005) to meet this standard:

1. Consider alternative sources of values for interpreting evaluation findings.
2. Provide a clear, defensible basis for value judgments.
3. Determine the appropriate party(ies) to make the evaluational interpretations.
4. Identify pertinent societal needs.
5. Identify pertinent customer needs.
6. Reference pertinent laws.
7. Reference, as appropriate, the relevant institutional mission.
8. Reference the programme's goals.
9. Take into account the stakeholders' values.
10. As appropriate, present alternative interpretations based on conflicting but credible value bases.

1.5 Report Clarity (US5): "Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes and findings of the evaluation, so that, essential information is provided and easily understood" (Sanders, 1994, p.24). Stufflebeam (2005) identifies the criteria as below for meeting this standard:

1. Clearly report the essential information.
2. Issue brief, simple and direct reports.
3. Focus reports on contracted questions.
4. Describe the programme and its context.
5. Describe the evaluation's purposes, procedures and findings.
6. Support conclusions and recommendations.
7. Avoid reporting technical jargon.
8. Report in the language(s) of stakeholders.
9. Provide an executive summary.
10. Provide a technical report (p.2).

1.6 Report Timeliness and Dissemination (US6): "Significant interim findings and evaluation reports should be disseminated to intended users, so that; they can be used in a timely fashion" (Sanders, 1994, p.24). The following criteria for meeting this standard are described by Stufflebeam (2005) as follows:

1. Make timely interim reports to intended users.
2. Deliver the final report when it is needed.
3. Have timely exchanges with the programme's policy board.
4. Have timely exchanges with the programme's staff.
5. Have timely exchanges with the programme's customers.
6. Have timely exchanges with the public media.
7. Have timely exchanges with the full range of right-to-know audiences.
8. Employ effective media for reaching and informing the different audiences.
9. Keep the presentations appropriately brief.
10. Use examples to help audiences to relate the findings to practical situations.

1.7 Evaluation Impact (US7): "Evaluations should be planned, conducted and reported in ways that encourage follow-through by stakeholders, so that, the likelihood that the evaluation will be used is increased" (Sander 1994, p.24). Stufflebeam (2005) outlines the following criteria for meeting this standard:

1. Maintain contact with audience.
2. Involve stakeholders throughout the evaluation.
3. Encourage and support stakeholders' use of the findings.
4. Show stakeholders how they might use the findings in their work.
5. Forecast and address potential uses of findings.
6. Provide interim reports.
7. Make sure that reports are open, frank and concrete.
8. Supplement written reports with ongoing oral communication.
9. Conduct feedback workshops to go over and apply findings.
10. Make arrangements to provide follow-up assistance in interpreting and applying the findings.

2 Feasibility Standards (FS1-FS3): "The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic and frugal" (Sanders, 1994, p.63) based on the following three standards:

2.1 Practical procedures (FS1): "The evaluation procedures should be practical to keep disruption to a minimum while needed information is obtained" (Sanders, 1994, p.63). Stufflebeam (2005) identifies the following criteria for meeting this standard:

1. Tailor methods and instruments to information requirements.
2. Minimise disruption.
3. Minimise the data burden.
4. Appoint competent staff.
5. Train staff.
6. Choose procedures that the staff are qualified to carry out.
7. Choose procedures in light of known constraints.
8. Make a realistic schedule.
9. Engage locals to help to conduct the evaluation.
10. As appropriate, make evaluation procedures a part of routine events.

2.2 Political Viability (FS2): "The evaluation should be planned with anticipation of the differing positions of various interest groups, so that, their cooperation may be obtained and, so that, possible attempts by any of these groups to curtail evaluation operations or bias or misapply the results can be averted or counteracted" (Sanders, 1994, p.63). Stufflebeam (2005) outlined the criteria for meeting this standard:

1. Anticipate differing positions of different interest groups.
2. Avert or counteract attempts to bias or misapply the findings.
3. Foster cooperation.
4. Involve stakeholders throughout the evaluation.
5. Agree on editorial and dissemination authority.
6. Issue interim reports.
7. Report divergent views.
8. Report to right-to-know audiences.
9. Employ a firm public contract.
10. Terminate any corrupted evaluation.

2.3 Cost-effectiveness (FS3): "The evaluation should be efficient and produce information of sufficient value, so that, the resources expended can be justified" (Sanders, 1994, p.63). Stufflebeam (2005) identified the criteria for meeting this standard:

1. Be efficient.
2. Make use of in-kind services.
3. Produce information worth the investment.
4. Inform decisions.
5. Foster programme improvement.
6. Provide accountability information.
7. Generate new insights.
8. Help spread effective practices.
9. Minimise disruptions.
10. Minimise time demands on programme personnel.

3 Propriety Standards (PS1-PS8): "The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically and with due regard for the welfare of those involved in the evaluation, as well as, those affected by its results" (Sanders, 1994, p.81). There are eight standards that have to be complied with under this category:

3.1 Service Orientation (PS1): "Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants" (Sanders, 1994, p.81). Stufflebeam (2005) outlined the criteria for meeting this standard:

1. Assess needs of the programme's customers.
2. Assess programme outcomes against targeted customers' assessed needs.
3. Help to ensure that the full range of rightful programme beneficiaries is served.
4. Promote excellent service.
5. Make the evaluation's service orientation clear to stakeholders.

6. Identify programme strengths on which to build.
7. Identify programme weaknesses to correct.
8. Give interim feedback for programme improvement.
9. Expose harmful practices.
10. Inform all right-to-know audiences of the programme's positive and negative outcomes.

3.2 Formal Agreements (PS2): "Obligations of the formal parties to an evaluation should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it" (Sanders, 1994, p.81). Stufflebeam (2005) identifies the following criteria for meeting this standard:
1. Evaluation purpose and questions.
2. Audiences.
3. Evaluation reports.
4. Editing.
5. Release of reports.
6. Evaluation procedures and schedule.
7. Confidentiality/anonymity of data.
8. Evaluation staff.
9. Metaevaluation.
10. Evaluation resources.

3.3 Rights of Human Subjects (PS3): "Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects" (Sanders, 1994, p.81). Stufflebeam (2005) identified the criteria for meeting this standard:
1. Make clear to stakeholders that the evaluation will respect and protect the rights of human subjects.
2. Clarify intended uses of the evaluation.
3. Keep stakeholders informed.
4. Follow due process.
5. Uphold civil rights.
6. Understand participant values.
7. Respect diversity.
8. Follow protocol.
9. Honour confidentiality/anonymity agreements.
10. Do no harm.

3.4 Human Interactions (PS4): Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that, participants are not threatened or harmed."(Sanders, 1994, p.81). Stufflebeam (2005 identified the criteria for meeting this standard:
1. Consistently relate to all stakeholders in a professional manner.
2. Maintain effective communication with stakeholders.
3. Follow the institution's protocol.
4. Minimise disruption.
5. Honour participants' privacy rights.
6. Honour time commitments.

7. Be alert to, and address, participants' concerns about the evaluation.
8. Be sensitive to participants' diversity of values and cultural differences.
9. Be even-handed in addressing different stakeholders.
10. Do not ignore, or help to cover up, any participant's incompetence, unethical behaviour, fraud, waste or abuse.

3.5 Complete and Fair Assessment (PS5): The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that, strengths can be built upon and problem areas addressed." (Sanders, 1994, p.82). Stufflebeam (2005) clarified the criteria for meeting this standard:

1. Assess and report the programme's strengths.
2. Assess and report the programme's weaknesses.
3. Report on intended outcomes.
4. Report on unintended outcomes.
5. Give a thorough account of the evaluation's process.
6. As appropriate, show how the programme's strengths could be used to overcome its weaknesses.
7. Have the draft report reviewed.
8. Appropriately address criticisms of the draft report.
9. Acknowledge the final report's limitations.
10. Estimate and report the effects of the evaluation's limitations on the overall judgment of the programme.

3.6 Disclosure of findings (PS6): "The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results" (Sanders, 1994, p.82). Stufflebeam (2005) provides the criteria for meeting this standard:

1. Define the right-to-know audiences.
2. Establish a contractual basis for complying with right-to-know requirements.
3. Inform the audiences of the evaluation's purposes and projected reports.
4. Report all findings in writing.
5. Report relevant points of view of both supporters and critics of the programme.
6. Report balanced, informed conclusions and recommendations.
7. Show the basis for the conclusions and recommendations.
8. Disclose the evaluation's limitations.
9. In reporting, adhere strictly to a code of directness, openness and completeness.
10. Ensure that reports reach their audiences.

3.7 Conflict of interest (PS7): "Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results" (Sanders, 1994, p.82). Stufflebeam (2005) clarified the following criteria for meeting this standard:

1. Identify potential conflicts of interest early in the evaluation.
2. Provide written, contractual safeguards against identified conflicts of interest.
3. Engage multiple evaluators.
4. Maintain evaluation records for independent review.

5. As appropriate, engage independent parties to assess the evaluation for its susceptibility or corruption by conflicts of interest.
6. When appropriate, release evaluation procedures, data and reports for public review.
7. Contract with the funding authority rather than the funded programme.
8. Have internal evaluators report directly to the chief executive officer.
9. Report equitably to all right-to-know audiences.
10. Engage uniquely qualified persons to participate in the evaluation, even if they have a potential conflict of interest; but take steps to counteract the conflict.

3.8 Fiscal responsibility (PS8): "The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that, expenditures are accounted for and appropriate." (Sanders, 1994, p.82). Stufflebeam (2005) outlined the following criteria for meeting this standard:

1. Specify and budget for expense items in advance.
2. Keep the budget sufficiently flexible to permit appropriate reallocations to strengthen the evaluation.
3. Obtain appropriate approval for necessary budgetary modifications.
4. Assign responsibility for managing the evaluation finances.
5. Maintain accurate records of sources of funding and expenditure.
6. Maintain adequate personnel records concerning job allocations and time spent on the job.
7. Employ comparison shopping for evaluation materials.
8. Employ comparison contract bidding.
9. Be frugal in expending evaluation resources.
10. As appropriate, include an expenditure summary as part of the public evaluation report.

4 Accuracy Standards (AS1-AS12): "Accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about features that determine worth or merit of the program being evaluated" (Sander 1994, p.125) based on the 12 standards:

4.1 Programme Documentation (AS1): "The program being evaluated should be described and documented clearly and accurately, so that, the program is clearly identified" (Sanders, 1994, p.125). Stufflebeam (2005) outlined the criteria for meeting the above standard:

1. Collect descriptions of the intended programme from various written sources.
2. Collect descriptions of the intended programme from the client and various stakeholders.
3. Describe how the programme was intended to function.
4. Maintain records from various sources of how the programme operated.
5. As feasible, engage independent observers to describe the programme's actual operations.
6. Describe how the programme actually functioned.
7. Analyse discrepancies between the various descriptions of how the programme was intended to function.

8. Analyse discrepancies between how the programme was intended to operate and how it actually operated.
9. Ask the client and various stakeholders to assess the accuracy of recorded descriptions of both the intended, and the actual, programme.
10. Produce a technical report that documents the programme's operations.

4.2 Context Analysis (AS2): "The context in which the program exists should be examined in enough detail; so that it's likely influences on the program can be identified" (Sanders, 2005, p.125). Stufflebeam clarified the following criteria for meeting this standard:

1. Use multiple sources of information to describe the programme's context.
2. Describe the context's technical, social, political, organisational and economic features.
3. Maintain a log of unusual circumstances.
4. Record instances in which individuals or groups intentionally or otherwise interfered with the programme.
5. Record instances in which individuals or groups intentionally, or otherwise, gave special assistance to the programme.
6. Analyse how the programme's context is similar to, or different from, contexts where the programme might be adopted.
7. Report those contextual influences that appeared to influence the programme significantly and that might be of interest to potential adopters.
8. Estimate effects of context on programme outcomes.
9. Identify and describe any critical competitors to this programme that functioned at the same time and in the programme's environment.
10. Describe how people in the programme's general area perceived the programme's existence, importance and quality.

4.3 Describe Purposes and Procedures (AS3): "The purposes and procedures of the evaluation should be monitored and described in enough detail, so that, they can be identified and assessed" (Sanders, 1994, p.125). Stufflebeam (2005) identified the following criteria to meet the above standard as below:

1. At the evaluation's outset, record the client's purposes for the evaluation.
2. Monitor and describe stakeholders' intended uses of evaluation findings.
3. Monitor and describe how the evaluation's purposes stay the same or change over time.
4. Identify and assess points of agreement and disagreement among stakeholders regarding the valuation's purposes.
5. As appropriate, update evaluation procedures to accommodate changes in the evaluation's purposes.
6. Record the actual evaluation procedures, as implemented.
7. When interpreting findings, take into account the different stakeholders' intended uses of the evaluation.
8. When interpreting findings, take into account the extent to which the intended procedures were effectively executed.
9. Describe the evaluation's purposes and procedures in the summary and full-length evaluation reports.

10. As feasible, engage independent evaluators to monitor and evaluate the evaluation's purposes and procedures.

4.4 Defensible Information Sources (AS4): "The sources of information used in a program evaluation should be described in enough detail, so that, the adequacy of information can be assessed" (Sanders, 1994, p125). Stufflebeam (2005) described the criteria for meeting this standard:

1. Obtain information from a variety of sources.
2. Use pertinent, previously-collected information once validated.
3. As appropriate, employ a variety of data collection methods.
4. Document and report information sources.
5. Document, justify and report the criteria and methods used to select information sources.
6. For each source, define the population.
7. For each population, as appropriate, define any employed sample.
8. Document, justify and report the means used to obtain information from each source.
9. Include data collection instruments in a technical appendix to the evaluation report.
10. Document and report any biasing features in the obtained information.

4.5 Valid Information (AS5): "The information gathering procedures should be chosen or developed and then implemented so that they will ensure that the interpretation arrived at is valid for the intended use" (Sanders, 1994, p.126). Stufflebeam (2005) identified the following criteria to meet the above standard as below:

1 Focus the evaluation on key questions.
2 As appropriate, employ multiple measures to address each question.
3 Provide a detailed description of the constructs and behaviours about which information will be acquired.
4 Assess and report the type of information each employed procedure acquires.
5 Train and calibrate the data collectors.
6 Document and report the data collection conditions and process.
7 Document how information from each procedure was scored, analysed and interpreted.
8 Report and justify inferences singly and in combination.
9 Assess and report the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions.
10 Establish meaningful categories of information by identifying regular and recurrent themes in information collected, using qualitative assessment procedures.

4.6 Reliable Information (AS6): "The information gathering procedures should be chosen or developed and then implemented, so that, they will ensure that the information obtained is sufficiently reliable for the intended use" (Sanders, 1994, p.126). Stufflebeam (2005) described the following criteria to meet the above standard:

1. Identify and justify the type(s) and extent of reliability claimed.
2. For each employed data collection device, specify the unit of analysis.

3. As feasible, choose measuring devices that in the past have shown acceptable levels of reliability for their intended uses.
4. In reporting the reliability of an instrument, assess and report the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions and the evaluator's bias.
5. Check and report the consistency of scoring, categorisation and coding.
6. Train and calibrate scorers and analysts to produce consistent results.
7. Pilot test new instruments in order to identify and control sources of error 8. As appropriate, engage and check the consistency between multiple observers.
9. Acknowledge reliability problems in the final report.
10. Estimate and report the effects of unreliability in the data on the overall judgment of the programme.

4.7 Systematic Information (AS7): "The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected" (Sanders, 1994, p.126). Stufflebeam (2005) clarified the following criteria to meet the above standard:
1. Establish protocols for quality control of the evaluation information.
2. Train the evaluation staff to adhere to the data protocols.
3. Systematically check the accuracy of scoring and coding.
4. When feasible, use multiple evaluators and check the consistency of their work.
5. Verify data entry.
6. Proofread and verify data tables generated from computer output or other means.
7. Systematise and control storage of the evaluation information.
8. Define those who will have access to the evaluation information.
9. Strictly control access to the evaluation information according to established protocol.
10. Have data providers verify the data they submitted.

4.8 Analysis of Quantitative Information (AS8): "Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered" (Sanders, 1994, p.126). Stufflebeam (2005) identified the following criteria to meet the above standard:
1. Begin by conducting preliminary exploratory analyses to ensure the data's correctness and to gain a greater understanding of the data.
2. Choose procedures appropriate for the evaluation questions and nature of the data.
3. For each procedure specify how its key assumptions are being met.
4. Report limitations of each analytic procedure, including failure to meet assumptions.
5. Employ multiple analytic procedures to check on consistency and replicability of findings.
6. Examine variability as well as central tendencies.
7. Identify and examine outliers and verify their correctness.
8. Identify and analyse statistical interactions.
9. Assess statistical significance and practical significance.
10. Use visual displays to clarify the presentation and interpretation of statistical results.

4.9 Analysis of Qualitative Information (AS9): “Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered” (Sanders, 1994, p.126). Stufflebeam (2005) clarified the following criteria to meet the above standard:

1. Focus on key questions.
2. Define the boundaries of information to be used.
3. Obtain information keyed to the important evaluation questions.
4. Verify the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders.
5. Choose analytic procedures and methods of summarisation that are appropriate to the evaluation questions and that employed qualitative information.
6. Derive a set of categories that is sufficient to document, illuminate and respond to the evaluation questions.
7. Test the derived categories for reliability and validity.
8. Classify the obtained information into the validated analysis categories.
9. Derive conclusions and recommendations and demonstrate their meaningfulness.
10. Report limitations of the referenced information, analyses and inferences.

4.10 Justified Conclusions (AS10): “The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them” (Sanders, 1994, p.126). Stufflebeam (2005) identified the following criteria to meet the above standard:

1. Focus conclusions directly on the evaluation questions.
2. Accurately reflect the evaluation procedures and findings.
3. Limit conclusions to the applicable time periods, contexts, purposes and activities.
4. Cite the information that supports each conclusion.
5. Identify and report the programme’s side effects.
6. Report plausible alternative explanations of the findings.
7. Explain why rival explanations were rejected.
8. Warn against making common misinterpretations.
9. Obtain and address the results of a pre-release review of the draft evaluation report.
10. Report the evaluation’s limitations.

4.11 Impartial reporting (AS11): “Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that, evaluation reports fairly reflect the valuation findings” (Sanders, 1994, p.126). Stufflebeam (2005) outlined the following criteria to meet the above standard:

1. Engage the client to determine steps to ensure fair, impartial reports.
2. Establish appropriate editorial authority.
3. Determine right-to-know audiences.
4. Establish and follow appropriate plans for releasing findings to all right-to-know audiences.
5. Safeguard reports from deliberate or inadvertent distortions.
6. Report perspectives of all stakeholder groups.
7. Report alternative plausible conclusions.
8. Obtain outside audits of reports.
9. Describe steps taken to control bias.

10. Participate in public presentations of the findings to help to guard against and to correct distortions by other interested parties.

4.12 Metaevaluation (AS12): “The evaluation itself should be formative and summatively evaluated against these and other pertinent standards, so that, its conduct is appropriately guided and on completion, stakeholders can closely examine its strength and weaknesses” (Sanders, 1994, p.126). Stufflebeam (2005) identified the following criteria to meet the above standard:

1. Designate or define the standards to be used in judging the evaluation.
2. Assign to someone the responsibility for documenting and assessing the evaluation process and products.
3. Employ both formative and summative Metaevaluation.
4. Budget appropriately and sufficiently for conducting the Metaevaluation.
5. Record the full range of information needed to judge the evaluation against the stipulated standards.
6. As feasible, contract for an independent Metaevaluation.
7. Determine and record the audiences, which will receive the Metaevaluation report.
8. Evaluate the instrumentation, data collection, data handling, coding and analysis against the relevant standards.
9. Evaluate the evaluation’s involvement of, and communication of, findings to stakeholders against the relevant standards.
10. Maintain a record of all Metaevaluation steps, information and analyses.

APPENDIX F

Summary of General Findings from Document Analysis

Table F1. Summary of General Findings from Document Analysis

Standards	Evaluation Case 1	Evaluation Case 2	Evaluation Case 3	Evaluation Case 4
Utility Standards (US1-US7)	**Fair**	**Good**	**Fair**	**Fair**
Stakeholder Identification (US1)	**Poor**	**Good**	**Good**	**Poor**
Evaluator Credibility (US2)	Fair	Very Good	Fair	Poor
Information Scope and Selection (US3)	Good	Fair	Fair	Good
Values Identification (US4)	Fair	Very Good	Very Good	Fair
Report Clarity (US5)	Excellent	Excellent	Excellent	Excellent
Report Timeliness and Dissemination (US6)	Very Good	Very Good	Good	Very Good
Evaluation Impact (US7)	Fair	Very Good	Very Good	Fair
Feasibility Standards (FS1-FS3)	**Good**	**Very Good**	**Good**	**Good**
Practical Procedures (FS1)	Very Good	Very Good	Very Good	Very Good
Political Viability (FS2)	Fair	Good	Good	Fair
Cost Effectiveness (FS3)	Very Good	Very Good	Good	Good
Propriety Standards (PS1-PS8)	**Good**	**Good**	**Good**	**Good**
Service orientation (PS1)	**Good**	**Very Good**	**Very Good**	**Very Good**
Formal agreements (PS2)	Very Good	Very Good	Good	Very Good
Rights of Human Subjects (PS3)	Good	Very Good	Very Good	Good
Human Interactions (PS4)	Good	Very Good	Good	Good
Complete and Fair Assessment (PS5)	Good	Good	Good	Good
Disclosure of Findings (PS6)	Good	Fair	fair	Fair
Conflict of Interest (PS7)	Fair	Fair	Fair	Fair
Fiscal Responsibility (PS8)	Very Good	Good	Good	Good

Table F1: (Continued).

Standards	Evaluation Case 1	Evaluation Case 2	Evaluation Case 3	Evaluation Case 4
Accuracy Standards (AS1-AS12)	**Fair**	**Fair**	**Fair**	**Fair**
Program Documentation (AS1)	Very Good	Very Good	Very Good	Very Good
Context Analysis (AS2)	Fair	Good	Good	Fair
Describes and Procedures (AS3)	Fair	Fair	Fair	Fair
Defensible Information Sources (AS4)	Poor	Good	Fair	Poor
Valid Information (AS5)	**Fair**	**Good**	**Good**	**Fair**
Reliable Information (AS6)	Fair	Poor	Poor	Fair
Systematic Information (AS7)	Good	Fair	Fair	Good
Analysis of Quantitative Information (AS8)	Fair	Poor	Poor	Fair
Analysis of Qualitative Information (AS9)	Fair	Fair	Fair	Fair
Justified Conclusions (AS10)	**Fair**	**Fair**	**Fair**	**Fair**
Impartial Reporting (AS11)	**Fair**	**Fair**	**Fair**	**Fair**
Metaevaluation (AS12)	Poor	Poor	Poor	Poor

Source: Ratings calculated from metaevaluation checklist based on Stufflebeam, (1999, p.2) and Tables A1, B1, C1 and D1 on p.196-205

APPENDIX G.

Summary of Findings on Compulsory Evaluation Standards

Table G1: Summary of Findings on Compulsory Evaluation Standards

Associate Standards	Evaluation Case 1	Evaluation Case 2	Evaluation Case 3	Evaluation Case 4
Service Orientation (PS1)	Good	Very Good	Very Good	Very Good
Valid Information (AS5)	Fair	Good	Good	Fair
Justified Conclusion (AS10)	Fair	Fair	Fair	Fair
Impartial Reporting (AS11)	Fair	Fair	Fair	Fair
Stakeholder Identification (US1)	**Poor**	**Good**	**Good**	**Poor**

Source: Ratings calculated from metaevaluation checklist based on Stufflebeam, (1999, p.2) and Tables A1, B1, C1 and D1 on p.196-205

APPENDIX H

Descriptive Statistics Findings

Table H1: Descriptive Statistics Findings on Utility Standards

	US1	**US2**	**US3**	**US4**	**US5**	**US6**	**US7**
Valid	4	4	4	4	4	4	4
Mode	**2**	**3**	**5**	**4**	**9**	**7**	**4**

Source: Own calculations based on ratings from Table F1 and Table G1 on p.225 and p.227

Table H2: Descriptive Statistics Findings on Feasibility Standards

	FS3	FS2	FS1
Valid	4	4	4
Mode	**7**	**4**	**8**

Source: Own calculations based on ratings from Table F1 and Table G1 on p.225 and p.227

Table H3: Descriptive Statistics Findings on Propriety Standards

	PS1	PS2	PS3	PSC4	PS5	PS6	PS7	PS8
Valid	4	4	4	4	4	4	4	4
Mode	**7**	**8**	**6**	**5**	**5**	**5**	**4**	**7**

Source: Own calculations based on ratings from Table F1 and Table G1 on p.225 and p.227

Table H4: Descriptive Statistics Findings on Accuracy Standards

	AS1	AS2	AS3	AS4	AS5	AS6	AS7	AS8	AS9	AS8	AS11	AS12
Valid	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
Mode	**7[a]**	**5**	**3**	**2**	**3**	**4**	**2[a]**	**4**	**4**	**3[a]**	**3**	**2**

Source: Own calculations based on ratings from Table F1 and Table G1 on p.225 and p.227