

**A Doctoral Thesis**

**"Metaevaluation of Programme Evaluation  
Practice in Kenya"**

**by**

**UNIVERSITY OF NAIROBI  
EASTAFRICANA COLLECTION**

**@**

**Karen T. Odhiambo**

**University of Nairobi**

**2003**

University of NAIROBI Library



0479186 9

# Metaevaluation of Programme Evaluation

## Practice in Kenya

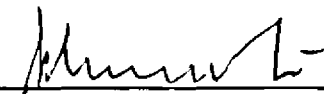
A research presented in fulfillment of the  
requirements for the degree of  
**Doctor of Philosophy**  
in  
**Measurement and Evaluation**  
Psychology Department  
of  
**The University of Nairobi**

---

### Supervisors:

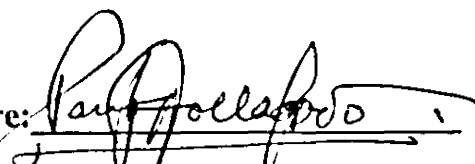
**Dr. Ernest L. Shivutse**

Psychology Department  
University of Nairobi

Signature:   
10/12/2003

**Prof. Paul W. Achola**

Sociology Department  
Kenyatta University

Signature:   
16/11/2003

## ACKNOWLEDGEMENT

I would like to acknowledge the professional guidance provided by my chief supervisor, *Dr. Ernest L. Shivutse* and second supervisor, *Prof. Paul W. Achola* throughout the research period as well as to those who contributed towards this task during data collection. I would like to also thank the staff of the *Kenya Institute of Education*, specifically, the *Research and Evaluation* department, without whom this work would not have been possible.

My appreciation also goes to the Canadian International Development Agency (CIDA) and the Kenya Government, through the Office of the President, for their financial support during my entire post-graduate education.

I cannot forget Nicolus Oriko, a friend to my late brother, Mody Ochako. To Nick, who continually motivated me and waited anxiously for this moment, I declare that by this acknowledgement to you, that I have officially completed the race making the doctorate official.

To my sister, Grace Ochako Omollo, who helped to produce the many reviews and final copies of this thesis, I say                    — *“Thank You Grele.”*

## DEDICATION

This thesis is dedicated to my family who stood by me throughout the time I spent on my studies, to my husband, David Odhiambo Ouma; to my daughters, *nyi-gem kowil*, that is, Brenda Anyango Keziah Oringo, Lorraine Atieno Phoebe Tocho, and little Daisy Adhiambo Kobongos' (Flora).

I also dedicate this work to my birth parents, Jared W. Ochako K'Adede and Phoebe Tocho Achieng K'Odero Ochako who have supported my education endeavors right from pre-school to university. Special dedication goes to my late parents of *Gem Sirembe*, Mzee Eliazaro Ouma, and Mama Keziah Oringo Ouma who gave me all the encouragement and support while they were still alive. God bless them all.

## **DECLARATION**

**I do hereby declare that this doctoral thesis research is my original work and has not been presented for a degree award in any other university.**

## ABSTRACT

This research sought to examine programme evaluation practice in Kenya. This was done by carrying out metaevaluation, that is, evaluation of evaluations that have been carried out on national programmes in Kenya. Metaevaluation is in itself an evaluation process which means that the research is based on the function of evaluation.

The impetus of the research was due to the fact that much as evaluation literature is full of discussions on philosophical and theoretical orientations to evaluation, very little has been done or written on evaluation practice as it actually takes place particularly in large settings, more so, within the African context. Also, evaluation has not always lived up to its own noble aspirations as noted by many social programmes worldwide. The researcher therefore thought that it was important to check the quality of the evaluation system in Kenya in order to establish how much is known of the process and to determine areas where better practice is needed. It was also hoped that the research would act as a catalyst to other similar researches.

The researcher looked into programme evaluation practice by researching on evaluation as practised in education, focusing on curriculum evaluation in schools in Kenya. The researcher chose this area because curriculum in education is one area where the Kenya government has commissioned large national evaluations making it possible to look at programme evaluation practice in diverse aspects.

The researcher analysed evaluation practice using “The *Programme Evaluation Standards*” developed by the Joint Committee (1994) as the criteria of performance to determine the acknowledged theory of evaluation practice, as well as an interview guide to determine the context in which evaluation takes place. The research design used was naturalistic inquiry by applying the audit trail content analysis and constant comparative method.

The results reveal lack of many desirable qualities related to specific principles of evaluation practice as prescribed by the *Standards*. Notably, defective principles were utility, serving information needs of intended users; feasibility, being realistic and politically viable; and propriety ethical standards, desiring protection of the rights of individuals. The accuracy standards, that is, technical aspects related to social science research approaches were better addressed. The results also reveal an evaluation context whereby the policy is not mediated or adapted, but, mandated and sanctioned at moments that suit interested parties.

One therefore, finds a situation that reflects a practice that is guided much more by political and technical aspects of evaluation but less of evaluation principles and methodologies. The research therefore concludes that programme evaluation practice in Kenya does not meet the *standards* criteria which are based on the acknowledged theory of evaluation. Further, the context does not support or facilitate the evaluation process. The results show that in order for an evaluation task to be completed, an evaluator needs to possess not only social science technical aspects of evaluation but also knowledge of all

the three programme evaluation *standards*, that is, utility, propriety and feasibility to comprehend the evaluation practice.

A critical analysis of the outcome of the evaluation research shows that, the Kenyan evaluation process presents a situation that is complex. The situation presents unresolved issues that need to be regarded as possibly new forms of evaluation approaches in a unique context, and that the findings may not always be a result of inappropriate practices but are determined by the situation at hand. The range of practice needed to meet the *standards* criteria also seems to be out of reach of an evaluator as theoretical evaluation issues that influence practice have not been addressed in the evaluation literature to guide practice, nor does the Joint Committee (1994) address these issues.



## TABLE OF CONTENTS

<b>CHAPTER ONE : INTRODUCTION</b>	<b>1</b>
1.1 Background of the study	1
1.2 Statement of the Problem	5
1.3 The purpose of the Study	6
1.4 Objectives of the Study	8
1.5 Research Questions	8
1.6 Justification of the Study	9
1.7 Significance of the Study	11
1.8 Limitations of the Study	12
1.9 Definition of Terms	12
<b>CHAPTER TWO : LITERATURE REVIEW of EVALUATION</b>	<b>16</b>
2.1 The Concept of Evaluation	16
2.2 Metaevaluation Process	19
2.3 The Concept of Evaluation	23
2.4 Definition of Evaluation	25
2.5 Trends and Theoretical Developments of Evaluation Literature	26
2.5.1 Early Influence : Stone age to 1965	27
2.5.2 Problems of the first three generations	37
2.5.3 Coming of age : The 1980s	39
2.5.4 The paradigm debate : Philosophy and Ideology	45

2.5.5	Approaches or models of the field	49
2.5.6	New Trends and Development : The 1990s	51
2.5.7	Overview and reflection of the evaluation literature	53

## **CHAPTER THREE : METHODOLOGY 57**

3.1	Research Framework	57
3.2	Feasibility of the Research Task	60
3.3	Data Collection Instruments	62
3.3.1	The Joint Committee of <i>Standards</i>	62
3.3.2	Semi-structured face-to-face interview schedule	65
3.4	Research Design	66
3.4.1	Content analysis	66
3.4.2	Hermeneutic dialectic procedure	66
3.4.3	Trustworthiness of the research : Validity and Reliability	67
3.5	Data Analysis	68
3.5.1	Audit trail analysis	68
3.6	Sampling procedure	70
3.6.1	Respondents for the study	72
3.7	Data Analysis	73
3.8	Preliminary Preparation of the Research	75

## **CHAPTER FOUR : RESULTS AND DISCUSSIONS 77**

4.1	Evaluation context	77
4.1.1	Policies that govern evaluation in Kenya	77

4.1.2	Establishing the Evaluation Mandate	82
4.1.3	Purposes of evaluation	87
4.1.4	Objects evaluated	89
4.1.5	Who conducts evaluation	91
4.1.6	Who calls for evaluation and funds evaluation	93
4.1.7	Determining preparation for evaluation	94
4.1.8	Research procedures followed	96
4.2	Theoretical Basis of Programme Evaluation Practice	97
4.2.1	Descriptive analysis of the Utility _ <i>Standards</i>	97
4.2.2	Descriptive analysis of the Feasibility _ <i>Standards</i>	130
4.2.3	Descriptive analysis of Propriety _ <i>Standards</i>	140
4.2.4	Descriptive analysis of Accurcy _ <i>Standards</i>	153
<b>CHAPTER FIVE : SUMMARY AND CONCLUSIONS</b>		<b>174</b>
5.1	Summary of the Metaevaluation Research Results	174
5.2	Synopsis of the Metaevaluation Outcome	194
5.3	Conclusion	198
5.4	Recommendations	201
<b>REFERENCES</b>		<b>205</b>
<b>APPENDICES</b>		<b>220</b>
Appendix _ A : The Joint Committee of standards		220
Appendix _ B : Format for Applying the Standards		223

Appendix _ C : Interview Schedules : Supplementary Information	234
Appendix _ D : Written Consent Form	240
Appendix _ E : Contact Sheet	242
Appendix _ F : The Research Data Base	245

## TABLES

Table: 2.5.5_1	Classification in terms of organisers	50
Table: 2.5.5_2	Classification in terms of belief system	51
Table: 3.1.1	Judgmental scale	59
Table: 3.6.1_1	Ministry of Education Officers (M.O.E) interviewed	72
Table: 3.6.1_2	Kenya Institute of Education(K.I.E) Officials interviewed	72
Table: 3.6.1_3	Donor organisations interviewed	73
Table: 4.1.5_1	Typical composition of evaluators	93
Table: 4.2.1_1	Commonly occurring purpose of evaluation	102
Table: 4.2.1_2	Objects of Evaluation	103
Table: 4.2.1_3	Quality and characteristics of statement of purpose	105
Table: 4.2.1_4	Internal evaluations : Descriptions and interpretations	108
Table: 4.2.1_5	External evaluations : Descriptions and interpretations	109
Table: 4.2.1.6a	Evaluation report format	112
Table: 4.2.1.6b	Evaluation document cover page	112
Table: 4.2.1_7	Schedule for field work	118
Table: 4.2.1_8	Decision taken based on the recommendations	123
Table: 4.2.1_9	Nature of statement of recommendations made	124
Table: 4.2.1_10	Quality of statements of recommendations made	125
Table: 4.2.1_11	Users and targets of evaluations	126
Table: 4.2.1_12	Orientation of evaluation	127
Table: 4.2.1_13	Sources of criteria for judgement	128

Table: 4.2.2_1	Instruments used in evaluations	131
Table: 4.2.3_1	Positive and Negative presentation of outcome	148
Table: 4.2.4_1	Samples of commonly arising objectives	158
Table: 4.2.4_2	Sources of information	159
Table: 4.2.4_3	Instruments used to solicit information	160
Table: 4.2.4_4	Units of analysis	161
Table: 4.2.4_5	Extent to which outcomes address objectives	163
Table: 4.3.1a	Summary of metaevaluation results	171
Table: 4.3.1b	Summary of standards achievement	173

## FIGURES

Figure 2.2.1	Metaevaluation Process	21
Figure 2.2.2	Evaluation process	22
Figure 3.1.2	Illustration of Research Framework	59
Figure 3.3.1.1	Joint Committee of Standards Criteria(JCS)	64

## ABBREVIATIONS

B.O.G:	Board Of Governors
C.I.P.P:	Context-Input-Process-Product model
C.P.U:	Central Planning Unit
D.E.P.P:	Division of Education Policy and Programme
D.F.I.D:	Britain's Department for International Development(DFID)
E.F.A:	Education For All
I.Cs':	Industrialised Countries
I.P.A.R:	Institute of Policy Analysis and Research
J.C.S:	Joint Committee of Standards
K.I.E:	Kenya Institute of Education
K.N.E.C:	Kenya National Examination Council
K.N.U.T.:	Kenya National Union of Teachers
L.I.C.s':	Least Industrialised Countries
M.O.E :	Ministry of Education, Science and Technology
N.D.P:	National Development Plan
N.G.O:	Non Governmental Organisations
N.M.E.C:	National Monitoring and Evaluation Committee
S.D.B.M.D:	Director of Budgetary Supplies Department and Budgetary Monitoring



T.A.C:	Technical Advisory Center
T.I.Q.E.T:	Integrated Quality Education and Training
T.S.C:	Teachers Service Commission
T.T.C:	Teacher Training Colleges
R.B.M:	Results Based Management
UNICEF:	United Nations Childrens' Fund

# CHAPTER ONE

## INTRODUCTION

In modern public organisations, it is rare not to find evaluation as an essential ingredient of any sector, whether properly applied or not. In both *Industrialised Countries* (ICs) and *Least Industrialised Countries* (LDCs) and at both national and local levels, evaluators are engaged in developing and testing innovative initiatives designed to improve and control social problems. They do this by refining and assessing the effectiveness and efficiency of programmes designed to improve the lives of people and alleviate social problems of the society. The success of these efforts depends on the quality of evaluation.

### 1.1 Background of the Study

An evaluation system needs to be checked for the quality of its output in order to establish how much we know of the process and to determine areas where better practice is needed. However, the evaluation community as a whole has not devoted much time to issues of quality or nature of evaluation process until the last decade (Forss and Carlsson, 1997). The concern about the quality of evaluation has led to interest in formal metaevaluation process.

Scriven (1994) states that the general discipline of evaluation has many specific and autonomously developed and applied areas. Among the well-known types are programme evaluation, personnel evaluation, performance evaluation, product evaluation, proposal

evaluation and policy evaluation. There are two other applied areas of great importance, that is, metaevaluation and discipline specific evaluation. The former is the *evaluation of evaluations*, the essence of this study, while the latter is the kind of evaluation that goes on inside an academic discipline.

Metaevaluation refers to the very function and practice of evaluation. The process of metaevaluation is therefore in itself an evaluation process. "Evaluation is the social process of making judgements of worth, a process basic to all forms of social behaviour whether that of a single individual, complex organisation, or an object with the aim of making decisions," (Suchman, 1967). Evaluation of utility is intrinsically interwoven with the development of knowledge.

Metaevaluation has existed as long as evaluation. Its purpose is to help evaluation live up to its potential. Evaluating evaluations helps promote quality in evaluation practice. According to Stufflebeam (1981), an evaluator's main concern is with appropriate evaluation practices, which are generally decided with some set *standards* for guiding and judging the evaluations. This is a process that has direct bearing on the quality of evaluation. He further states that the concern of good evaluations requires that evaluation enterprise itself be evaluated. Properly conducted and practised evaluations lead to desired improvement of performance and operations of programmes. There is need to carry out metaevaluation to

determine performance of the evaluation processes used in all social settings in order to ascertain the effectiveness of the evaluation process. A good evaluation should contain criteria by which its effectiveness may be determined. The primary concern of the evaluator is to generate information basically for decision making. A wrong decision as a result of an inefficient evaluation system is likely to lead to adverse effects on those who rely on the decisions made.

Evaluation is a psychological metaphor essentially viewed as basic form of human behaviour (Worthen & Sanders, 1987). To contribute to psychological function is not only accomplished by studying individual human behaviour but also direct investigation of happenings between organisational operations. These include those of stakeholders, programme managers, beneficiaries and politicians; individual operations and influences such as those of policy makers and evaluators; evaluators and organisational approaches such as methodology, decision making process as well as communicating and utilising evaluation results. Psychological inquiry of programme evaluation involves operations of all aspects of the programme under investigation and focuses on outcomes that extend our knowledge to all aspects of the programme. Bootzin et. al. (1991) states that the term accountability equated with evaluation, has brought many fields into evaluation including psychology since it addresses performance of programmes.

House (1993) states that evaluation has not always lived up to its own noble aspirations. This is reflected in the magnitude of increasing economic crisis world wide despite evaluation

being part and parcel of these programmes. Most social programmes, education included, have failed to meet their objectives of improving human life, yet they have been monitored and evaluated. The situation is worse in Africa (Asamoah, 1988). This may mean that evaluation process as applied may not have met the standards of appropriate evaluation practice. Unfortunately, there is lack of empirically based knowledge on the practice of programme evaluation. Relatively, little is known about evaluation as it takes place in practice. Statements of this nature are found across the literature of evaluation (Smith, 1983; Wilcox, 1989; House, 1993).

House further states that much as evaluation literature is full of discussions of philosophical and theoretical orientations of evaluation, very little has been written on evaluation practices. Wilcox (1989), further, states that useful as the philosophical and theoretical orientations may be, relatively little has been written about evaluation practices as it takes place particularly in large settings. According to Asamoah (1988), the situation is worse in LICs especially in Africa, as very little has been done in terms of research related to empirical studies in evaluation and that evaluation is under-utilised, depriving the policy maker of empirical data on which to base social decisions.

There is need to look into this inadequacy within the African context so as to boost literature on evaluation practice in LICs and in Africa in particular. The researcher has

chosen to do this by looking into evaluation practices in Kenya.

## **1.2 Statement of the Problem**

Literature available in Kenya reveals lack of dependable information system on programme evaluation with no long standing tradition of utilising empirical data derived from knowledge produced by evaluation. This has deprived the country of dependable information needed for decision making and policy development. The researcher has observed that where evaluations have been carried out, the interest in the results draws more attention than the evaluation process itself. Also, politics seems to play a central role in influencing programme evaluation process, especially large ones, such as those in education.

There is evidence of failures of programmes, which reflects a need for systematic evaluation research. Where evaluation practice has been studied, issues raised relate to planning, refining purposes, procedures and priorities of evaluation; questions concerning use of results; no clear definition of what constitutes an evaluation activity; the value put on evaluation operations, as well as consequences of the programmes (Conner, 1985; Clemson, 1985, Mulusa, 1988). Knowledge gained from such studies could be useful in improving evaluation practice generally, especially in Kenya.

A study by Wilcox (1989) and another by Lynch (1988) exemplified this problem and the

authors agreed that what is really needed for the improvement of evaluation and its practice is to publicly verify evidence of applications of evaluation methods, particularly in various settings. In order to contribute to theory and practice, evaluation research must be grounded within programme site and must deal with issues pertinent to practitioners. There is therefore need for research that asks significant questions concerning certain issues, such as, the kinds of objects evaluated, the purpose of evaluation, the methods used in conducting the evaluation, the people who participate in the evaluation and factors associated with utilisation of evaluation results.

A critical investigation of the literature does not reveal an indication of a comprehensive research of this nature on programme evaluation practices in Kenya. New approaches to societal reforms and the central place that programme evaluation holds today open new avenues for research into evaluation practices and consequently, metaevaluation.

### **1.3 The Purpose of the Study**

The purpose of the study was to examine programme evaluation practice in Kenya by carrying out a metaevaluation. The study looked into programme evaluation practice by studying programme evaluation processes in education. This was narrowed down to the curriculum in primary and secondary schools.

The researcher did this by describing and critically, analysing, the results and making recommendations that would prove helpful in the development of acceptable evaluation policies and practices.

The researcher chose to look into the area of programme evaluation practice in education since this is one area where there have been evaluations of large programmes commissioned by the government and also because of the recognition of education as a cornerstone of economic and social development and reform efforts. This gave the study diverse aspects which were helpful in bringing out heterogeneous levels of evaluation operations and procedures involving many categories of individuals and groups within a complex context. Evaluation here involved all purposes of evaluation discipline in general and applied full range of evaluation activities that required programme needs in terms of the purpose for evaluation, technical needs in terms of social science methods and evaluation methodological approaches. All this was based on agreed upon model ranging from initial programme planning, object of inquiry, as well as interpreting and utilising the evaluation results.

The **rationale** of this study was that, evaluation based on systematic programme evaluation practices would play an important role in upgrading the practice of evaluation, thereby improving the operations and performance of evaluations conducted in Kenya.



## 1.4 Objectives of the Study

The following evaluations were identified to guide the research. The objectives of the study were designed to:

- i) Examine the *context* in which evaluation takes place
- ii) Determine the *purpose* of evaluation
- iii) Identify programme evaluation *policies* in Kenya
- iv) Determine the *content* of evaluation, and, *entities* that are evaluated
- v) Determine whether evaluation practice meets the *theoretically acknowledged body of knowledge* on evaluation practice by authorities in evaluation in terms of utilisation of evaluation results; feasibility and practicality of evaluations; the welfare of participants and technical adequacy of the results.

## 1.5 Research Questions

The following research questions were designed as a guide to help judge the adequacy of the evaluation process. To meet the objectives of the study, the research questions addressed the *context of evaluation* and evaluation practiced as prescribed in *acknowledged theoretical evaluation body of knowledge* as defined by the Joint Committee of Standards (1994) based on four criteria for measurement, that is, *utility, feasibility, propriety* and *accuracy*, thus reflecting the characteristics of appropriate programme evaluation practices. The research questions are listed below:

- i) What are the government's programme evaluation *policies* ?
- ii) What is the *mandate* in respect to evaluation?
- iii) What *purposes* is programme evaluation meant to meet?
- iv) What particular *objects* are evaluated?
- v) *Who* conducts evaluation and in *what* capacity?
- vi) *Who calls* for the evaluation or *funds* evaluation?
- vii) To what extent do the evaluations serve the *information needs* of the stakeholders, utilisation and dissemination of the results\_ *Utility*
- viii) To what extent are the results *feasible* in terms of practical procedures, organisational behaviour and what are the constraints\_ *Feasibility*
- ix) To what extent are the evaluations conducted with regard to the *welfare of the stakeholders*, that is, are they complete and fair assessment, disclosure of findings and conflict of interest\_ *Propriety*
- x) To what extent do the evaluations *technically* reveal adequate information about features that determine worth or merit of the programme being evaluated\_ *Accuracy*

## 1.6 Justification of the Study

During the past two decades, the need for systematic research in programme evaluation has been emphasised by social scientists and policy makers in both ICs and LICs. The need is particularly acute as LICs struggle to design social programmes that will be useful for the majority within the constraints imposed by fragile economics and limited resources.

Asamoah (1988) states that in LICs information on evaluation process is not made readily accessible as findings are usually confidential and are often intended for internal use. Also, evaluation activities do not have a long standing tradition such as the one in the United States of utilising empirical data for decision making and policy development. He further states that funds provided for evaluation in LICs are generally scarce and mandatory evaluation based on legislation requiring evaluation such as practised in the developed world, are few. Smith et. al. (1993) also observes that even where evaluation has been inevitable, it did not create enough interest among administrators to set *standards* for the conduct of evaluation or to assess evaluation quality. Some programmes such as those in education at the time of independence in LICs or those designed and implemented after independence do not match with current socio-economic realities. Evident failures of social-economic programmes despite massive inputs, further attest to the need for systematic evaluation research on evaluation practice (Worthen and Sanders, 1987).

Asamoah (1988) further states that the situation is worse in Africa where evaluation is underutilised and very little has been done in terms of research related to empirical studies of this nature. This deprives policy makers of empirical data on which to base evaluation decisions. Empirical studies that exist have focused on studies done for particular research purposes. Very few have looked at other aspects of evaluation processes, such as methodology and dissemination of results. Also, most of the studies that exist on the practice of programme evaluation have been reflections by researchers from LICs on their experience

resulting from carrying out evaluations in LICs. This further attests to the need for research on programme evaluation process by indigenous people.

## **1.7 Significance of the Study**

A detailed empirical examination of the evaluation process could potentially enhance both evaluation research and evaluation practice. It could enhance evaluation research by providing data on which hypotheses might be generated, for example, questions about utilisation of results, the views of the stakeholders involved in evaluation and influence of organisational behaviour. Research of this nature contributes to practice in that knowledge of performance may challenge and change the way practitioners think about problems that arise while conducting evaluations and the way they carry out their tasks. Practitioners would benefit from introspection of their actions and asking questions about the way they carry out their tasks, thus increasing their knowledge of evaluation practice.

The findings of this study will be useful to administrators who are required to make decisions about evaluation policies and procedures. The results will also contribute to literature on programme evaluation practice in Kenya and serve as reference material for scholars and practitioners. Training needs for evaluators could also be identified and training programmes developed. Most studies of this nature have been conducted in ICs. This study on metaevaluation in Kenya will certainly create local culture for evaluation.

## **1.8 Limitations of the Study**

The study was carried out in Kenya on programme evaluation practices of curriculum programmes at primary and secondary school levels, thus limiting the scope to which the research results may be applied. The researcher relied on information from the documents that were made available by the curriculum development centre, the Kenya Institute of Education(K.I.E) which was also the location of the study, and had no control over documents made available. The researcher also relied on the reflections provided by the respondents from interviews and had no control over the respondents' reflections on the topic under investigation. A social science study of this nature is likely to encounter problems of both sampling and instrument limitations as the study cannot cover the entire population and the instruments cannot elicit all required information.

## **1.9 Operation of Terms**

### **Accountability:**

Responsibility to account by showing that project implementation is proper and that the project is reaching intended targets and that there are cost benefits.

### **Evaluation:**

It is the determination of the merit and worth of an entity with a view to making

decisions. In education, it is the formal determination of the quality, effectiveness or the value of a programme, product, project, process, objective, or a curriculum using systematic inquiry and judgement methods.

**Evaluation Practice:**

Way or mode of doing things, in this case evaluation.

**Evaluand:**

The entity to be evaluated.

**Evaluation standards:**

A principle mutually agreed to by people engaged in the professional practice of evaluation that if met, will enhance the quality and fairness of an evaluation.

**Evaluator:**

Any person who conducts an evaluation.

**Evaluation context:**

The environment within which evaluation takes place, for example, socio-political environment; policies that govern evaluation.

**Least Industrialised Countries (LICs):**

These are countries, that when compared to **Industrialised Countries (ICs)**, have substantially low gross national product, per capita income, employment rate, life expectancy, health status, literacy rate, quality communication and media with a large percentage of population living below the defined poverty level. They are also characterised by high birth rates.

**Metaevaluation:**

Evaluation of Evaluation

**Performance:**

An accomplishment or achievement in accordance to prescribed criteria.

**Programme:**

This term has been used to refer to the object of evaluation related to programme activities that are provided on a continuing basis which includes content, projects, materials and activities.

**Programme Evaluation:**

It includes all purposes of evaluation in general as applied to the full range of evaluation activities from initial design, planning, process, determining merit and worth, legal and technical adequacy, utilisation and communication of results.

**Stakeholders:**

Individuals or groups that may be involved in or affected by a programme and evaluation outcome (not including the client, that is, the government)

**Standard:**

Criteria by which a process is assessed

**Values:**

Established ways of doing things that members of a given profession regard as desirable, things that make sense only in terms of some structure and form.

**Worth:**

Real merit that is useful and of importance to what comes to be or comes about and happens justly, rightly, suitably, deservedly in a timely manner and effort.



## CHAPTER TWO

### LITERATURE REVIEW

In this chapter, literature related to the study has been reviewed under the following sub-themes: the concept of metaevaluation, metaevaluation process, concept of evaluation, definition of evaluation, trends and theoretical development of evaluation literature.

#### **2.1 The Concept of Metaevaluation**

The concept of metaevaluation is derived from the fact that not all evaluation activities are valuable, or well-intentioned. Questions have been raised as to whether the results obtained from evaluations warrant their time and cost or not. According to Nilsson & Hogben (1983), practitioners and advocates of metaevaluation interpret the discipline as *evaluation of evaluation*. The practitioners main concern is with good evaluation. This could mean evaluation of evaluation *itself* and evaluation of particular specimens of evaluation where evaluation is the determination of value or worth. When one talks of valid procedures, justified decisions and recommendations, one is actually engaged in making value judgements. In other words evaluations can be appraised as evaluations just as we appraise science as science. According to the two authors, if these arguments can be said to be intelligibly right or wrong (correct or incorrect), then it can be stated that evaluations are based on theory, founded within “evaluation literature.”

The on-going discussions are based on the function of evaluation itself. In other words, *metaevaluation* refers to the very function and practice of evaluation, and that, evaluation does lead to uncontested improvements that would not occur in any other way. Nilsson & Hogben (1983) further state that, due to many failures of evaluation and the fact that evaluation has not lived up to its expectations, it is understandable why many more question the concept of evaluation. Worthen and Sanders (1987) argue that, failure in evaluation cannot however be blamed on its concept but in the way in which evaluation is conducted. Evaluation in itself is a worthwhile activity. One cannot however say with certainty that all evaluation activities are intrinsically valuable or even done with good intentions. This has resulted into statements that look into whether the results warrant the costs in terms of human resource and time, and whether the activities are appropriate in producing credible results, in generating and utilising evaluation results, and in aiding decision making.

The purpose of metaevaluation is, therefore, to help evaluation live up to its standard professional expectation. Scriven (1991), notes that metaevaluation represents ethical and scientific obligations when the welfare of others are involved. That is, metaevaluation is professional imperative. So, when one asks questions to determine whether an evaluation is good, the suggestion is that this should be decided with a set of standards or criteria internationally acknowledged for guiding and judging evaluations (Stuffelbeam, 1981).

Metaevaluation has existed as long as evaluation. Informally, it has dealt with having an opinion about the quality of evaluations. During the 1960s, evaluators began to discuss formal metaevaluation procedures and criteria leading to the idea of a set of standards for guiding and judging evaluations. Unpublished checklists of evaluation standards began to be exchanged informally among evaluation authorities. Several evaluators published their proposed guidelines or metaevaluation criteria for use in judging evaluation plans or reports. The 1980s also brought about the idea of unifying the developing field (Evaluation Research Society, 1980; Joint Committee of Standards, 1981; Stufflebeam et al., 1971; U.S General Accounting Office, 1978; Nevo, 1981). Much as some writers have criticised the rationale for the whole standard-setting effort as being premature at the present state of the art in evaluation, there is a great deal of agreement regarding their scope and content.

Various standards have come up in the literature over time. One of the most significant and the most elaborate and comprehensive set developed in 1981 was the *Standards for Evaluation of Education Programmes, Projects, and Materials* which has received widespread attention in evaluation (Joint Committee, 1981:1994; henceforth referred to as the *Standards* or *Joint Committee*). Worthen and Sanders (1987) state that the standards is the ultimate benchmark against which both evaluation and other sets of metaevaluation criteria and *standards* should be judged. The Joint Committee of Standards is the criteria that was used in this research. Other efforts are also identified in the literature. For example, The Evaluation Research Society (ERS) was designed to serve a wide range of applied social science disciplines and health, law enforcement and

public policy fields. Boruch and Cordray (1980) analysed the set of standards and reached the conclusion that there has been a great deal of overlap and similarity amongst them. However, evaluators accept the Joint Committee of Standards as the canon of practice in evaluation.

The *Joint Committee of Standards (JCS)* were designed by a seventeen-member Joint Committee of Standards for education evaluation as a result of a pioneer project started in 1975 to ensure an ethical approach to evaluation of educational programmes. It was developed by representatives of twelve professional groups involved in and concerned with the practice of evaluation. The purpose was to guide the professional practice of a growing, local, state and national enterprise in need of professional direction and control. The standards are not geared towards any evaluation model or theory but are organised around four important attributes of an evaluation, namely, utility, feasibility, propriety and accuracy.

## **2.2 Metaevaluation Process**

It has been established that the concept of metaevaluation is based on the function of evaluation itself. Therefore the process for metaevaluation is in itself an evaluation process. The difference in metaevaluation is determined by the procedures used in an evaluation that has been conducted and documented.

According to Payne (1994), there will probably never be a total agreement on the nature of the activities and sequence of steps in evaluation processes. The principles of the scientific procedures generally guide the practice. This involves identifying the area of inquiry, collecting data as designed, analysing and interpreting the results. The purpose of the evaluation, resources available, and time lines are some of the things that dictate the final form of the process. The sequence of activities may differ depending on the evaluation demands and also the methodological approach chosen by the evaluator (compare qualitative versus quantitative). The activities would depend on the roles of evaluation which could be summative, the extent to which the goals of the programme has been met, or formative, the extent to which the objectives of the programme as planned and being achieved. The nature of evaluation generally involves retrieval of information from documents as well as field studies. In an education programme, for example, the data collection process could range from simple interview with students to determine their attitudes towards schooling to surveys on students achievement across the republic.

Figure 2.2.1 and 2.2.2 illustrate an outline of the activities that generally define both the procedures of evaluation and metaevaluation. Each approach in the literature would require a unique model but all of them have a commonality that is related to the prescribed procedure.

In conducting metaevaluation, that is, in assessing the process described for evaluation in the preceding paragraph above, an important component is the specification of criteria or

*standard* settings needed for use in determining credibility of laid down value foundations laid down for the evaluation itself, or, making value judgements on the evaluators' performance based on the criteria.

**Figure: 2.2.1: Metaevaluation Process**

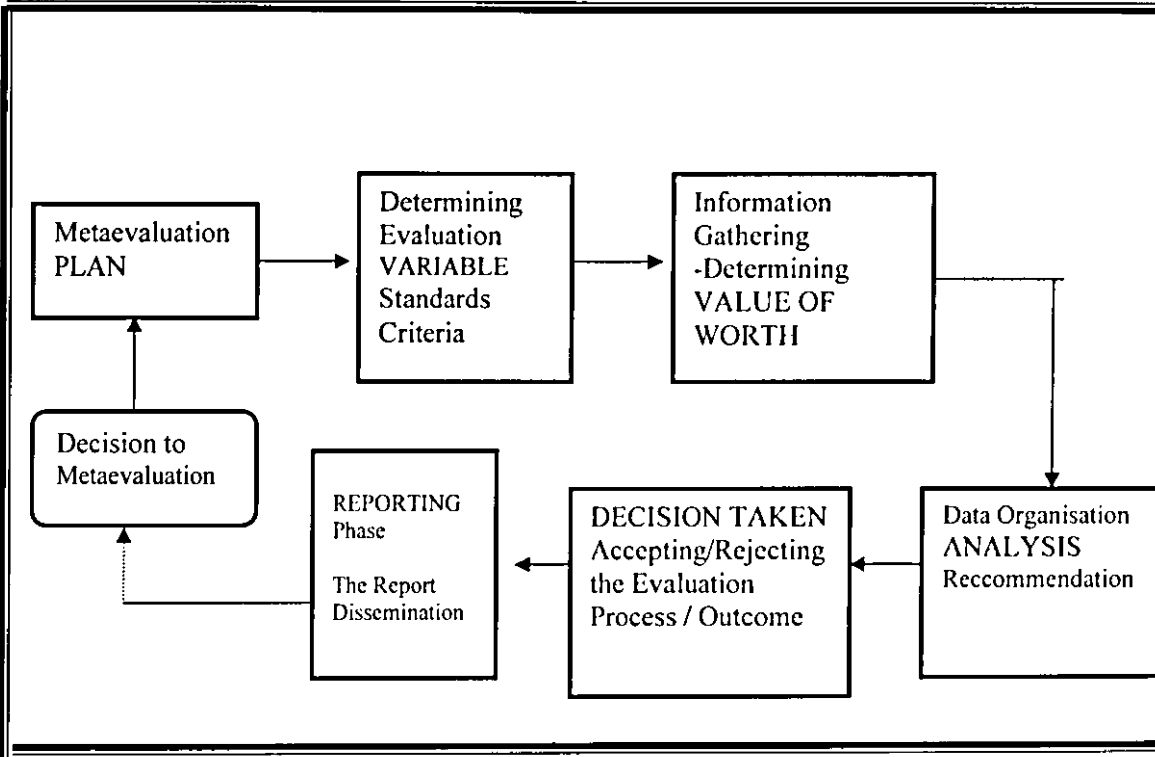
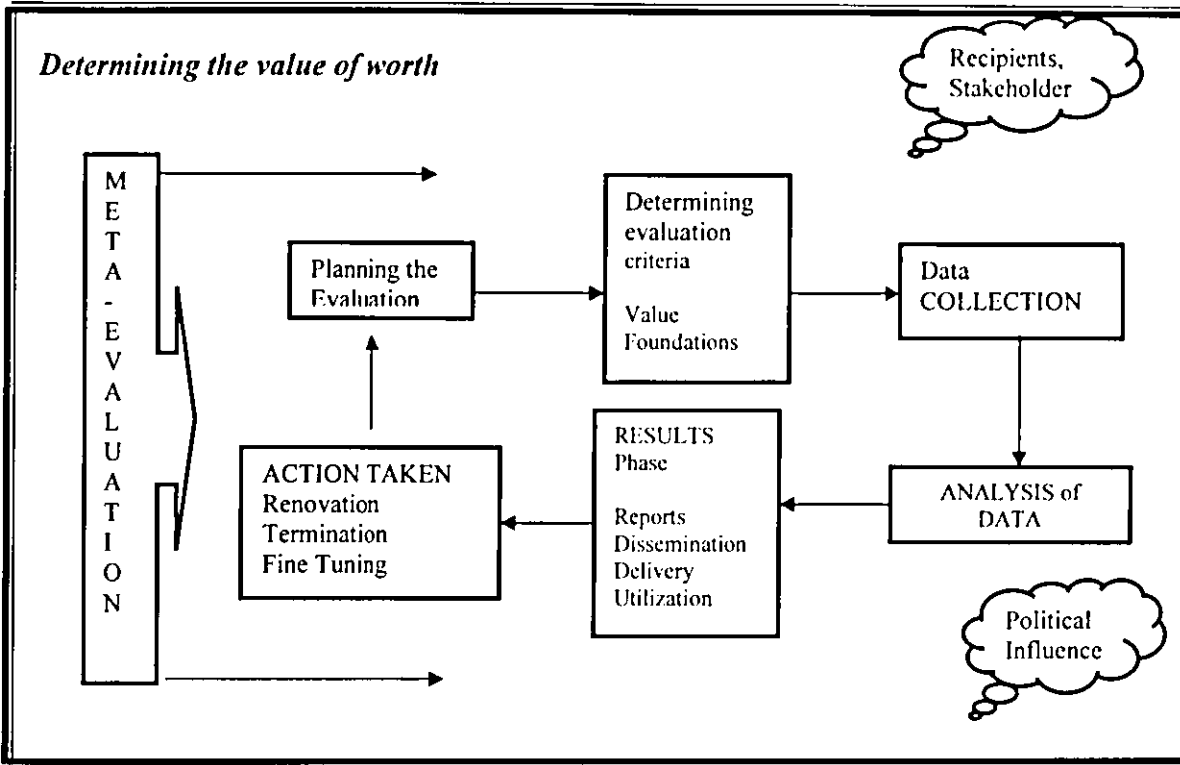


Figure: 2.2.2 : Evaluation Process



The concern for *standards* in the evaluation profession arises from the need to produce meaningful and defensible evaluations where the outcome or results are accurate and appropriately used. These are concerns of ethics and professional *standards* in the emerging evaluation profession. Values do permeate the *standards* that define credibility issues resulting in the *standard* criterion problem. At the very beginning when identifying objectives and goals that have evaluative priority, values are asserted. Also, while collecting data, judgements are continuously being made that address any emerging *discrepancy* between expected and observed outcomes. In other words, to establish the worth of an evaluation, it is necessary to carry out a metaevaluation.

## 2.3 The Concept of Evaluation

Evaluation has always been with us. It is a component of every activity and is generally used to distinguish between the best and less dependable alternatives involving matters related to practice, events, objects, processes, including the evaluation process itself.

In Bloom's *Taxonomy of Education Objectives* (Bloom, 1956), distinctions are made between six levels of hierarchically arranged cognitive operations which range from knowledge at the lowest level, to comprehension, followed by application, analysis, synthesis and lastly evaluation at the highest level. These hierarchies of cognitive taxonomies categorise the way in which information is used thus developing a scale ranging from simple concrete behaviour, through complex and more abstract behaviour such as evaluation (Popham, 1993). Evaluation which is at the highest abstract level is made about the value of methods, objects and persons for particular purposes and divided into judgements in terms of *internal* evidence and those of *external* criteria.

Scriven (1994) states that human beings evaluate hundreds of things every day, for example, the clothes we wear; the grilled beef as to how rare or well done; and the weather in terms of how warm or cold. No matter how informal the efforts made are to appraise the quality of something. Evaluation is an ancient practice being an integral part of every practice. The flint chippers, stone carvers and craft workers left testimony of their increasing evaluative knowledge of procedures that led to improvement of



indicators of poor performance. Every craft work uses evaluation as the process for quality control measure. Evaluation is also found in various fields such as academic disciplines, such as mathematics as a pure subject; transdisciplines such as statistics; engineering's quality control procedures; medical ethics; appellate court jurisprudence; the legal assessment of legal opinions; and in scholarly book reviews.

Evaluation can be done arbitrarily by wine experts and art critics or, it can be done conscientiously and accurately as in scoring achievement tests. This is to state that, evaluation can either be informal or formal. The former is generally based on subjective perceptions while the latter is based on systematic scientific approaches.

The types of evaluations to be discussed here are those based on systematic efforts of evaluation discipline as practised in evaluating social programmes, a process that permeates all areas of thought and practice whose task is the systematic and objective determination of merit, worth or value of an entity. Systematic assessment of practice of personnel and programme dates back to 2200 B.C. in China. This was when the then emperor instituted proficiency requirements for public officers demonstrated in formal tests (Guba and Lincoln, 1985).

## 2.4 Definition of Evaluation

Given such broad focus of the term *evaluation*, Worthen et, al. (1990) observe that, there is no uniformly agreed upon definition of precisely what the term evaluation means. Evaluation is not a new concept. Looking at some dictionary definitions of evaluation, Barnhart (1993) observes, “to evaluate is to find out the value or the amount of; to estimate the worth or importance of; and to appraise.”

The most preferred definition is that proposed by Scriven (1967), who defined evaluation as *determining the merit of worth of something*. They state that there are other definitions that arise in the literature such as: research or measurement; assessment of the extent to which objectives have been met; a professional judgement; auditing or several of the variant quality control; and as an act of collecting and providing information for intelligent decision making. The authors believe that evaluation is the *determination of the worth and merit of an evaluation object*. This is a definition accepted by most authorities. Evaluation becomes the identification, clarification, and application of defensible criteria to determine an evaluation object’s value or worth and merit, quality, utility, effectiveness, or significance in relation to those criteria. They further state that evaluation uses inquiry and judgement methods, such as: standards for judging; collecting relevant information; applying standards to determine value, quality, utility, effectiveness or significance.

Evaluation is generally equated with the number of terms such as measurement, appraisal research, and accountability that guided its inception and are used interchangeably (Worthen and Sanders (1987). Measurement is the quantitative description of behaviour, things or events. As a process for data collection, it is a key tool in research and evaluation. Engaging in research is not evaluation, but an attribute that is being studied, for example, achievement and attitudes in appraising learners for the purpose of diagnosis. Research, according to the two authors, is a systematic inquiry aimed at obtaining generalisable knowledge by testing claims about the relationships among variables, that the resulting knowledge may or may not have immediate application or implications as in evaluation. Accountability is a term equated much with evaluation, a term that has become an important driving force in the growing field. The continuing drive towards accountability has prevailed making evaluation an important social activity. The authors contend that one's definition is the product of what one believes is the purpose of evaluation.

## **2.5 Trends and Theoretical Development of Evaluation Literature**

Evaluation is troubled by ideological disputes just as it is by definition. Evaluations practised today have evolved through a number of forms. These can be traced over time and will be sketched briefly looking at the changed meanings of programme evaluation over time, and purposes that practitioners have perceived in mind for conducting evaluations. Guba and Lincoln (1981) acknowledged that over time the construction of

evaluation has become more informed and sophisticated despite the fact that there has been no consensus on any form of evaluation.

In every field of study beyond the basic practice is the step to formulate a guide to govern that practice. This is an investigative or practical procedure for improving our understanding of practice, that is, the methodology. In evaluation, most methodologies have been developed within the confines of the field of programme evaluation (Scriven, 1994). Programme evaluation literature has developed without an acceptable philosophy and has been influenced by two major paradigms, the positivists who apply quantitative data collection methods, and, the naturalists who apply qualitative data collection approaches. The basic set of beliefs held by these two paradigms have led to today's diversity in evaluation theoretical perceptions, resulting in numerous models or approaches of evaluation practice.

## **2.5.1 Early influence of programme evaluation: Stone age to 1965**

### **2.5.1.1 *First generation evaluators***

Formal evaluation can be traced back to 2000 B.C among the Chinese. Greek teachers such as Socrates also used verbal mediated evaluations as part of the learning process. Formal valuation of education and social programmes did not start until the mid-nineteenth century. Educators have been most influential theorists, an unusual accomplishment for researchers in the field of education (House, 1993). Education evaluators usually start their historical accounts with Ralph Tyler, an individual referred to as the father of programme evaluation and one also credited with the objective oriented

approach of evaluation. The objective oriented approach remains popular to-date despite its shortcomings.

Previous experience with testing and curriculum evaluation gave educators a headstart. Testing tradition has been one of the factors that has influenced the discipline. Measurement was applied to various attributes of school children, to evaluate the performance of students, and to determine mastery of curriculum content. Perhaps the earliest formal attempt to evaluate performance of schools took place in the United States – Boston in 1845. Previously, oral examinations were given to students. As the number of students increased, there was need to standardise the tests leading to the introduction of essay type of tests to make possible inter-school comparison. In terms of programme evaluation, there was a hidden policy agenda on decisions concerning annual appointments of head teachers. Madaus et. al. (1983) state that this was an example of politicisation of evaluation data, a phenomenon that has had far reaching influence in evaluation practice in LICs’.

Between 1887-1888, Joseph Mayer Ryce, an American, conducted what was considered to be the first formal evaluation in education. He wanted school time to be spent not only on the basics, but also to include art, music, and other subjects. He succeeded in sorting out the problem. More importantly to programme evaluation, was his argument that educators had to become experimentalists and quantitative thinkers and his use of comparative research to study children’s achievement. Ryce also proposed a system of resolving policy issues, an approach that came to be known as advocate-adversary or

judicial approach to evaluation. Another development was the foundation of the accreditation or professional judgement approach to evaluation.

Other applications to testing that had implications were in France by Alfred Binet a psychologist. He devised the first *Intelligence Test* which was later used to screen personnel for the army in the United States. The American version, the Stanford Binet Test, has become a permanent feature of the American system.

A second factor that played a role in the development of programme evaluation during the first generation of evaluation was the legitimisation provided by the enormous rise and application of *social science* to the study of human and social phenomena. The approach used here was the quantitative measurements that tests were yielding. Also, the scientific management in business and industry which emphasised systematisation and *efficiency*. The tests were seen as the means of determining whether pupils measured up to the specifications that the school had set leading to the coining of the term *measurement and evaluation*.

From the fore-going discussion, it can be seen that tests played a key role in unfolding the evaluation metaphor. Its influence led to proliferation in school tests during the 1920's and 1930s during which time *measurement and evaluation* came to be used interchangeably. The term evaluation was more often used to mean assigning grades to students or summary of their performance, a position that is still common to date. The term evaluation as it is known today was still evolving.

Guba and Lincoln (1989) state that this first generation of evaluation could be referred to as the *measurement generation*. The role of the evaluators was technical in that they were expected to have full knowledge of the entity to be measured and to be able to develop required the instrument. This technical sense of evaluation persists today as evidenced in education by the practice requiring pupils to pass tests as part of their high school certification, college admission, ranking of schools, determining the effectiveness of teachers and by publication of textbooks that use the phrase measurement and evaluation.

#### **2.5.1.2 *Second generation evaluators : The breakthrough***

This generation came about as a result of deficiency in the methods of the first generation who only targeted students as the object of evaluation. Due to influx of students from the Second World War torn countries, there was need to streamline the curriculum to make it relevant. The evaluation approach that provided only data on student performance could not serve the purpose for the kind of evaluation that was desired. It was realised that there was need to utilise students test results beyond individual performance towards programme effectiveness. However, the approaches applied were informal and critics were not impressed (Worthen and Sanders, 1987). This process was started as early as the 1930s and continued into the 1960s. It was only in the 1970s that evaluators realised this. The curriculum as evaluated was not generalisable and that it was not sensitive to students. The 1930s disillusionment led to the landmark Eight Year Study in the United States, launched in 1933, that resulted into a formal plan of evaluation that remains popular today(Guba and Lincoln, 1985).

Ralph Tyler was a member of the Bureau of Education Research at Ohio State University where the programme was launched. During this time, Tyler came to view evaluation not as an appraisal of students but rather as the appraisal of a programme, in this case curriculum programme. The purpose of the studies was to refine the developing of the curriculum and to make sure it was working or meeting the objectives set for it and that programmes needed to be judged to the extent that they promoted students mastery of the objectives that the programmers had established prior to their initiative. This approach went beyond mere student performance but also considered the curriculum output and impact. Thus, *PROGRAMME EVALUATION* was born. The resulting information from these evaluations led to the refinement of the curriculum, a process we would today refer to as *formative evaluation*.

The second generation evaluators were characterised by description of strength and weaknesses of programmes with respect to stated objectives. The role of the evaluators was that of a *describer* much as the earlier technical measurement aspects of the evaluators roles were also retained. Measurement was no longer viewed as equivalent to evaluation but as one of the tools of programme evaluation.



### 2.5.1.3 *Third generation evaluators : Coming of age*

This era has seen evaluation develop to high conceptualisation levels carrying with it debates in programme evaluation development and theory. These debates continue to date.

Evaluators in this era were stimulated by the focus on schools when the Americans were astonished by the first satellite launched by the Soviet Union in October 1957. Money was given to develop science and mathematics curricula. Subject specialists found little to go by to determine whether their instructional materials worked. The programme developers declined to commit themselves to the objectives approach of the Tylerian era until they had a clear picture of what they were doing. Their fear was that objectives would close off the creativity prematurely. Their argument was based on the fact that should an evaluation show deficiency at the final stages of programme development, it would be too late to do anything about it.

The issues arising in the on-going discussion are well documented in Cronbach's now classic *Course Improvement Through Evaluation*, a seminal article of 1963 in which he argued that if educational evaluation were to be of assistance to programmes, in this case the curriculum developers, it had to be focused on the decisions faced by the curriculum specialists during the process of their development efforts. The argument is that, programme evaluation activities should deal less with comparisons between programmes and more with the *degree* to which a given programme promotes its desired consequences. The article called for new directions on programme evaluation and made

suggestions to reconceptualise programme evaluation and review evaluation as a process for gathering useful data for revision and improvement of programmes. Specifically, he made three major points:

- that if evaluation was to be of maximum utility to the developers of new courses, it needed to focus on the decisions that those developers had to make *during* the time that development was occurring, known as *formative evaluation*.
- that evaluation needed to focus on the ways in which refinements and improvements could occur while the course was in process of development known as *monitoring*.
- that if evaluation were to be of maximum utility to the developers of new courses, it had to be more concerned with course performance characteristics than with comparative studies known as *impact*.

Issues of monitoring, formative evaluation and impact are important attributes in programme evaluation in relation to project development and management as well as programme performance. Although Cronbach's recommendations did not have an immediate impact, the article later helped spark a greatly expanded view of programme evaluation in the decade that followed and became a landmark to evaluation literature which began to reflect the increasing emphasis on the importance of programme evaluation.

Later, Cronbach's call for new directions was critiqued in Scriven's (1967) classical paper on *The Methodology of Evaluation*. According to Guba and Lincoln (1985) this was the single most important paper written on evaluation. Some of Scriven's contributions were involved: drawing a distinction between *formative* and *summative* evaluation or between

*improving* and *judging* the evaluand; a call for professional evaluators to take on themselves the burden of rendering judgement not just results; distinguishing evaluation from mere assessment of goal attainment as to whether the goals are in themselves worth achieving; distinguishing intrinsic or process evaluation from payoff or outcome evaluation; and contrasting the utility of comparative evaluations with that of non-comparative evaluation which was a direct opposition to Cronbach's utility of comparative evaluations.

Other evaluation authorities also offered their views as well, resulting in a proliferation of new evaluation alternatives. Some argued that statements of objectives were not value neutral but were based on certain implicit metaphors that guide thinking about the nature of the evaluation. As a result three major metaphors came up: the industrial metaphor from the era of scientific management; the behaviouristic metaphor from behavioural psychology; and the biological metaphor from developmental theories in biology. The statement of objectives rests heavily on the first two of these metaphors, while, for example, elements such as teachers are affected much more by the third metaphor. Reference was made to conventional objectives as *instructional* objectives, while those based on the biological metaphor were referred to as *expressive* objectives. It was further stated that *expressive* objectives cannot be dealt with in terms of a *common standard* and that, evaluation of these objectives required a reflection upon what had been produced in order to reveal its uniqueness and significance.

Guba (1969) pointed to certain clinical signs of evaluation's failures; lack of adequate definition of evaluation and lack of adequate theory; lack of knowledge of decision processes, lack of criteria on which to base judgements; lack of approaches to match the level of the complexity of the evaluand; lack of mechanism for organising, processing, and reporting evaluative information; and lack of trained personnel. According to Guba and Lincoln (1985), what was really needed was a fundamentally different approach to evaluation. This led to the constructivist view that employs qualitative measures advocated by the naturalistic paradigm.

The objective oriented approach which had served for two decades had some serious flaws which only came to light in the post-Sputnic period. These evaluators had neglected what Robert Stake(1967) called the other countenance or face of evaluation, that is, judgement. This marked the emergence of the third generation of evaluators.

Yet another impetus to development of programme evaluation was the evaluation mandate that the United States of America attached to the great society education legislation of 1965, a mandate that spreads evaluation to other social programmes and beyond. This was the mandate of Secondary School Act passed in 1965. This resulted in numerous programmes intended to overcome the injustices associated with poverty and race. Accountability became an issue beyond that of judgment. To ensure accountability of funds dished out, mandatory evaluations were mandated for monitoring the funds and accounting for the same (Guba and Lincoln, 1985). In the recent past, evaluation has been mandated by donor funds associated with aid to LICs bringing with it results-based-

management(RBM), a call that has seen monitoring and evaluation components institutionalised beyond individual programmes to national programme as a condition for donor funding.

As can be seen from the literature, despite the need for guidance on appropriate approaches to guide practice, there was little to refer to except better developed methodologies such as experimental design, psychometrics, curriculum development and survey research. Rarely, were the evaluations exemplary, thus revealing the conceptual and methodological impoverishment of evaluation in that era. The problem was further compounded by lack of trained evaluators. This situation became a stimulus for the rapid expansion of programme evaluation in the late 1960s. Several scholars in other specialisations also became interested in evaluation process. Stake (1967) wrote on the issue of judgement, the other countenance or face of evaluation. He stated that, without judgement, evaluation was not broad enough. Other writers also began to focus on other organisers. Examples of such writers were, Stufflebeam's Context-Input-Process-Product model (CIPP) and Alkin's Center for Study of Evaluation Model. Evaluators who take organiser's approach do not require objectives approach, but base their work on decisions to be made, the person to make the decisions, on what schedule, and by what criteria to make the decisions.

The call to include judgement in evaluation marked the *third generation evaluation*. Here the evaluator assumed the role of a judge and retained the earlier technical and

descriptive functions of evaluation as well. The contributions of evaluators greatly broadened the earlier views of evaluation.

### **2.5.2 Problems associated with the first three generations.**

Three pervasive problems which plagued these early generation evaluators were identified. These have been discussed at length by Guba and Lincoln (1989). They include:

#### **2.5.2.1 *Tendency towards managerialism:***

*Managerialism* denotes the clients or sponsors who commission or fund evaluation, the personnel responsible for implementing the evaluation report, for example, the policy maker, and manager of agency. In this study the manager was the Ministry of Education (MoE). This relationship can lead to a number of undesirable consequences such as: not being blamed for failures of evaluations; disempowering and unfair in that the manager has ultimate powers to determine what questions to ask, how the data are to be collected how the results are to be used and by whom; disempowering of the stakeholders who may have a difference of opinion on what other questions to ask and interpretations to make; disenfranchising as the manager retains the right to determine if the evaluation findings are to be released, and to whom, resulting in denied privilege of information hence the stakeholders rights; possibility of manager-evaluator relationship becoming cosy whereby evaluations done to gain the manager's approval is likely to lead to other contracts.

Guba and Lincoln (1989) state that while a vast majority of evaluators would not succumb to this tendency, the first three generation of evaluators could easily slip into such state of affairs.

#### **2.4.2.1 Over commitment to the scientific paradigm:**

This has to do with emulating methods of physical science with conviction and enthusiasm. This is the need to be rational and systematic in the physical science format. The problems associated are: assessment of the evaluand as though it did not exist in a context but only under controlled conditions; a situation referred to as *context stripping* which is not possible in the social world; over independence on formal quantitative measurement data that can be measured with precision, a position that assumes that what cannot be measured cannot be real, thus closing out any alternative approaches such as qualitative measurement data; the evaluator in this paradigm is relieved of any moral responsibilities towards the findings and the use to which the evaluation results may be put. The evaluator's work basically ends with the delivery of the report.

The issues discussed led to new approaches towards alternative modes of evaluation approaches. Despite many criticisms levelled at objectives-oriented approach, many evaluators continued to show the influence of Tyler. This is true to-date. Some of the evaluators have been Hammond cube (1973); Provus discrepancy model (1971); Popham's instructional objectives approach (1975); and Stake's countenance model (1976), probably the best known.

### 2.5.3 Coming of age the 1980s

Mention has been made of the idea to unify the evaluation developing field. This resulted in *standards* for evaluation practice. About the same time that the *standards* were being considered, a gradual change was taking place. This had to do with the realisation of the complexity of settings in which evaluation was taking place. As the field of evaluation expanded, many countries became incorporated in evaluation both the LICs and the ICs. Africa included. Many articles have addressed the uniqueness of evaluation in different countries and cultures within a country. It has been realised that what works in one setting /place may not work in another, and that, even within the same structure in one classroom, children are not the same. The children's socio-economic backgrounds play a role in their eventual performance in academics. The earlier belief that non-western objectives to western social science could be answered by a combination of great sensitivity to cultural differences and more attention to methodology has been discounted. It has been realised that concepts like control, predictability, generalisation may not be valued in contexts nor are they congruent with modes of thinking in western societies.

Writings on international and cross-cultural evaluation became noticeable in the literature of the late 1970s and early 1980s. Various concerns have been realised and their effects addressed. For example, and Cronbach (1982) wrote on cross-cultural evaluation and referred to evaluation as service-oriented, client focused process which takes place in particular institutional and political contexts that require a special kind of consideration. In many LIC settings, evaluation is basically an outside concern making it a form of neocolonialism as most of the developing countries were once colonies of foreign powers



such as Britain and France. As a result, formalised evaluations have largely been instigated by foreign needs and demands (Smith, 1993). This dates back to colonial times whereby, representatives of foreign powers were required to send back reports to their superiors. After independence, such reports were conducted by foreigners for the purpose of aid agreements with national or international institutions, such as Non-Governmental Organisations (NGOs) and United Nations (UN) organisations. Thus, one might say that formalised evaluation though existent in most LICs is largely a foreign concern and that evaluations mainly serve donor interests and rarely do they attend to national interests (Asamoah, 1988). This has led to questioning the practical utility of impact evaluations in these settings.

One reality of LICs which varies from country to country, is that of political expediency which strongly influences decisions about policy and planning, a situation that is likely to prevail with or without evidence that supports or fails to support the usefulness of a programme. According to Asamoah, (1988) researches having distinct policy implications that influence policy are treated with suspicion, and findings are not always put to practical use. However, there is no dispute on the necessity of evaluation.

In most LICs there is a dual economy model that specifies asymmetry of production and distribution conditions between two sectors, one *traditional* and the other *modern*. The traditional sector which is basically rural, is characterised by a paucity of capital, a poverty level of income of slow growth or stagnation. The modern sector, in contrast, is predominantly urban and more industrialised with modern communication systems and

infrastructure. It is wholly market oriented with modern commercial, financial and transport facilities, an innovation of the 20<sup>th</sup>. century technology. Wages are substantially higher than in the traditional rural sector, a trend likely to increase absolutely or relatively for generations given the normally high population growth in these countries. The system is also characterised by hierarchical system of governance (Taut,2000). These characteristics make the LIC world setting unique providing a challenge in terms of values to employ and methodological approaches to use in conducting evaluations (Child F.C., 1977; Brass, W. & Jolly, C.L., 1993).

The challenge of development in the latter part of the 20<sup>th</sup> century improvement of quality of life, especially in the poorer countries of the world, involves not just higher income but better education, better health and nutrition, more quality of opportunity, and a richer cultural life. Succeeding in development and sustaining it is a pressing challenge (World Report, 1991). As a result of this, the 1990's began with dramatic changes by the LICs initiating ambitious reforms of their economic and political system in trying to meet this challenge. There has been a surge of activity that has resulted in initiation of domestic and international foreign funded social programmes costing billions of dollars. The popularity that has given rise to increasing citizen pressure for more social development programmes that are bigger and better. The major focus of these programmes has been upon action, or the development of operational programmes and delivery of services. The demand for determining the effectiveness of such services and social action programmes has become increasingly high from funding agencies and the

government, local people, social scientists as well as evaluators, a demand that has been and still is growing everywhere (Suchman, 1967; Asamoah, Y.,1988; House, E., 1993;).

However, the *criteria* available for use in determining the *effectiveness* of social programmes in such a context remains unsolved, ( House, E., 1993; Shardish, W .R. & Reichardt, C.S., 1987; Shardish, W.R., Cook, T.D. & Leviton, L.C., 1991). The need is particularly acute as LICs struggle to design social developing programmes that will maximise positive benefits for the majority within the constraints imposed by fragile economies and limited resources as time runs out. The urgency is great and there is to arrest the situation before it reaches unmanageable proportions.

Evaluation also takes place within particular authority structures and cultures in the LICs. According to House (1993), evaluation in the United States government has been the blending of applied social science and economic decision making. As hundreds of evaluations are conducted by the federal government (at a cost of about 100 million dollars a year), the evaluations basically focus on the narrow goals of the programme rather than the broader issues and structures. The only exception are the independent critical evaluations. He concludes that evaluations therefore lend themselves to careers of high level bureaucrats who need to make their programmes look good.

Smith(1993), raises the question as to whose cultural values should dominate in the conduct of evaluation. According to Chapman & Boothroyd (1988), evaluation in LIC context does not lend itself to the use of traditional scientific methods. Cronbach (1982) cautions that the evaluators aim is not to diminish or control the effects in this context,

but to use the in-context information to *illuminate* the evaluation so that ultimately programmes can be improved. The knowledge of evaluation theory, research design, statistics and psychometrics is no longer sufficient as evaluators have to deal with the complexity of management, policy, value economics and psychological questions. The importance of contextual variables led to a conclusion that there was no one way to do evaluations (Patton, 1985; Joint Committee, 1981) and no inquiry into a particular programme at a particular time, with a particular budget (Cronbach, 1982).

Cronbach et. al.,(1982) argued earlier that because of the problem thus stated, evaluation should then be re-conceptualised. Their arguments are given in the form of ninety-five theses which cover issues such as qualitative-quantitative data gathering, flexibility of evaluation design, communication of evaluation results and characteristics of evaluation. They advocate flexibility in evaluation design and that evaluators should choose whatever approaches seem accurate considering practical and political considerations.

Patton summarises a number of these cross-cultural accounts and states, “each of these positions makes sense in the particular context in which it was made. They are context-bound non generalisations” (1985. p. 93). Patton notes over one hundred and fifty (150) definitions of culture related to political culture, bureaucratic culture, development culture, national culture, regional culture and culture of science. Ginsberg (1988), Chapman & Boothroyd (1988) concluded that epistemological and methodological issues also need attention. Smith (1993) contends that all this puts the evaluator in a context of competing values related to special interest groups and value positions. These positions which are more dramatic and complicated in cross-cultural studies which according to

Ginsberg (1988) may reflect not just differences of opinion, but differences of epistemology and moral philosophy. Ginsberg however criticises the position of writers like Connor and Patton on the ground that they advocate treating fundamental cultural differences like simple differences in U.S stakeholder opinions. Connors (1985) had earlier advocated more flexible studies, while Patton (1985) stressed the need for cultural sensitivity.

Smith, (1991) states that the dilemmas identified on whose values and criteria to choose are serious issues whose discussions are still noticeably absent in the literature on cross-cultural evaluation. He notes that the differences are likely to be related to differences in values, belief systems and notions of causality. The cultural differences should be seen as opportunity to enlarging evaluation's worldview and not as a back set. The field could stand to gain a great deal from these revelations. As LICs develop more of their evaluation capacity, non-western forms of evaluation are likely to emerge. LICs specifically Africa, provide an excellent laboratory for cross-cultural theory testing and refining cross-cultural research techniques.

Key questions arising here are concerned with the *standards* that should be invoked when working in a different culture. The question that arises, is whether such cultural practices are violations of the Joint Committee. This also raises the question as to the extent to which the Joint Committee or any standards criteria are *universal* or *culturally* biased. Many possible forms of criteria exist, but, the literature lacks consensus on what criteria to use. There is diversity and partiality of approaches and the presentation depends on

who is addressing the issue. There has been no systematic analysis of the conditions that can or should guide evaluation of these programmes in this context.

Attention to the practicality of practical and political issues and the brevity of their descriptions has opened evaluators' minds and brought to surface the apparent complexity that has accompanied the growth of evaluation literature. This situation has been described by Patton (1982) as one "filled with overflowing uncertainty, ambiguities, competing perspectives, and conflicting roles.' The emerging issues make the world of evaluators difficult. Making sense of the confusion has been the undertaking of a number of writers.

An area of interest in programme evaluation has been evaluation utilisation; hence questions such as: how are evaluations used? when are they used? why are they used? This has also seen the shift from the tendency for the field to focus on technical and methodological issues. Introduction of *utilisation* by Patton (1986) challenged the scientific attitude as a critical criterion for evaluation excellence. Patton states that, one weakness is that few theoretical process-models have been developed. This further calls for the need for additional research in the area.

#### **2.5.4 The paradigm debate: Philosophical and ideological influences**

The assumptions held by different perspectives about knowledge and value give rise to different evaluation practices in relation to methods of conducting evaluations. The problems of the first group of evaluators and the subsequent preceding development of

evaluation literature have demonstrated evaluations constructions, that is, the interpretations held, as well as its level of construction. These have greatly influenced evaluation as it is viewed today. Evaluation has developed without an accepted philosophy to form the theoretical knowledge base for evaluation literature. Indeed, the extensive discourse and debate concerning the philosophical assumptions of epistemology and value are largely responsible for the diversified view about evaluation practice.

Major flaws have been explored during the period of the first three generations. These relate to failure to accommodate value-pluralism, over commitment to the scientific methodology and tendency towards managerialism. Scriven (1994) contends that, the era continues to suffer from certain flaws which raise questions as to whether additional refinements or even a complete reconstruction is needed. The need to refine approaches is as a result of the underlying assumptions embedded in the belief of each evaluation approach. Worthen and Sanders (1987) state that, "there is no univocal philosophy of evaluation, any more than there is a single, universally accepted philosophies of knowing or establishing the truth (epistemology) and how it affects the approach to evaluation one would use." They hold the position that much of today's diversity in evaluation theoretical perceptions on methodology (also referred to as models or approaches by some theorists) of evaluation have been influenced mainly by two *paradigms*, the scientific convention method, positivistic approach (using quantitative data collection methods) and the naturalistic constructive approach (using qualitative data collection methods).

In their argument for the naturalistic constructivist view, Guba and Lincoln (1989) presented the following properties: that evaluation outcomes are constructions that individuals form to make sense in the situation they find themselves, that is, the reality (versus context stripping); that the world in which people find themselves is shaped by values more so in the context of LICs (versus value free); that constructions are subject to human error as they are a product of human thought, and that, hard data characteristic of quantitative measures are not feasible in the social world. Here, evaluations are shaped to empower stakeholders (versus disenfranchisement) by involving them in design, implementation, evaluation and to have a say in the utilisation of the results. This has to do with full *participative* and involvement, a phenomenon that has gained much attention today. Guba, however, states that, there is no way of addressing these issues in any unambiguous and certain way or in any way capable of proof, and that, the set of answers given in each case is the basic belief system of paradigms.

By 1985, qualitative naturalistic approaches were no longer shunned by most thoughtful evaluators and had become respected though the quantitative tradition still prevailed. The balance of power has however shifted rapidly towards qualitative approaches \_ perhaps too rapidly (House, 1993). The dialogue has begun to move beyond acrimonious debate between the two basic paradigms and evaluators have discovered the benefits of using both methodologies within a single evaluation. It is hoped that the next decade will see more ecumenical resolution of this issue. This debate is seen as a major cause of rift that permeates the field of evaluation and that, what is considered as acceptable evaluation, depends on the position taken regarding one or another of these contentions.



There are a number of alternatives that fall in between the two philosophical approaches. Guba suggests a more appropriate label, "*ideologically oriented inquiry*," including neo-marxism, materialism, feminism, freerism, participatory inquiry and critical theory itself. All the applications are tied together because they reject the claim of value freedom made by positivism, but hold the claim that paradigms are human constructions and so they inevitably reflect the values of their human constructions. They enter into inquiry at choice points and maintain that nature cannot be seen as it really is or as it really works, except through a *value* window.

The fact is that if evaluators confront a situation of value-pluralism, it must be the case that different views will emerge from persons and groups with differing value systems, such as the feminist movement. This approach is based on classical liberalism which sees society as an association of self-determining individuals who cooperate with others for self-interested ends. It is essential that they have a direct hand in governing themselves since they know themselves and their interests best. The sanctity of the individual against the intrusion of the society is paramount with the individuals being free to pursue their own goals. In the public sphere, the society has a right to impose some rules of equity or fairness. This brings up the issue of how to handle difficult disparate regions as found in LICs such as urban /rural disparities. This fact has also been discussed by many writers in the literature without coming up with concrete solutions. The question that arises is what an evaluator can do to ensure that evaluators are presented, clarified, and honoured in a balanced and even-handed way?

From the foregoing discussion, it is not easy to determine what values and whose values shall govern evaluation. According to Guba and Lincoln (1981), if the findings of studies can vary depending on the values chosen, then the choice of a particular value system tends to empower and enfranchise certain parties while disempowering and disenfranchising others. Inquiry thereby becomes a political act. However, none of these paradigms is the paradigm of choice. Each is an alternative that deserves to be considered on its own merit. The dialogue is not to point out this merit but rather to take practitioners and theorists to another level at which all of these paradigms will be replaced by yet other paradigms, and as the author states, the outlines we can reflect on now but dimly, if at all. It will not be a closer approximation of the truth, but will simply be more informed and sophisticated than those being entertained here.

### **2.5.5 Approaches or models of the field**

Collectively, the new conceptualisations of evaluation greatly broadened the earlier views held by evaluators. These new and controversial ideas fed the developing evaluation field with vocabulary and literature leading to a *plethora* of evaluation articles and books in the two decades, that is, the 1970s and 1980s. According to Worthen and Sanders (1987), this has resulted in at least forty formalised or semi-formalised evaluation approaches . These are generally organised into classifications in terms of what they take as organisers such as objectives and decisions as well as responsive evaluation or depending on the belief system of the approach.

**Table: 2.5.5\_1 : Classification in terms of organisers**

Approaches	Dimensions
<b>Objective Oriented</b>	The focus is specifying <i>goals and objectives</i> and determining the extent to which they have been attained
<b>Management Oriented</b>	The central issue is on identifying and meeting the information needs of <i>managerial decision makers</i> .
<b>Consumer Oriented</b>	The central issue is on developing evaluative information on " <i>products</i> ," broadly defined, for use by consumers in choosing among competing products, and services
<b>Expertise Oriented</b>	These depend primarily on the direct application of <i>professional expertise to judge</i> the quality of whatever endeavour is evaluated.
<b>Adversary Oriented</b>	<i>Opposition in points of view</i> of different evaluators (pro and con) is the central focus of the evaluation.
<b>Participatory Oriented</b>	This is where <i>involvement of participants</i> (stakeholders in that which is evaluated) is central in determining the values, criteria, needs, and data for evaluation.

Derived from Worthen and Sanders (1987)

**Table: 2.5.5\_2 : Classification in terms of a belief system**

Classification	Belief System
<b>The strong-decision support</b>	Attention is on whether programmes reach their goals, that is objective oriented, a view exemplified by Ralph Taylor.
<b>The weak-decision support</b>	They hold that decision support provides decisions relevant data but does not draw conclusions.
<b>The relativistic support</b>	The view is that evaluation should be done by using the clients values as a work frame without any judgement by the evaluator.
<b>The rich of thick descriptive approach</b>	Here evaluators report what they see without trying to make evaluative statement or infer to evaluative conclusions or even in terms of the clients values, for example, ethnographic or journal enterprise.
<b>The social process school</b>	Understanding social programmes with an aim of helping them improve, crystallised by a group of Stanford academics led Lee Chronbach.
<b>The Transdiscipline approach</b>	They reject the view that evaluation is a search for quality, merit and worth in favour of the idea that it is the result of construction by individuals and negotiation groups.

Derived from Scriven (1994)

### 2.5.6 New trends and developments : 1990

In the last decade of the twentieth century, there has been an upsurge of further emphasis on accountability-oriented environment. There has been great emphasis on poverty in the society yet there is an unwillingness to increase government funding for social problems (Mowbray et. al., 1998). Many professionals in the field speculate that this signals a new era in which evaluation will gain rise in prominence. Evaluation is continually being

made routine as an aspect of operations not only for state, but, also business and industry and professional associations. If evaluation is to sustain a resurgence and play a prominent role in policy and decision-making in the twenty-first century, the field may need to take a more realistic view towards its operations. Over the last decade, organisations and governments have been forced to make major changes in the way they develop and provide services and products. Out of this experience has come the notion of:

- **Empowerment**

This is one of the latest development in the field (Fetterman, 1994), an approach that uses evaluation concepts and techniques to foster self-determination and includes training, facilitating, advocacy, illumination and liberation.

- **Participatory**

This refers to engagement into joint action by actors, a term associated with literature on community development, a concept that is people centered approach to development (Turnbull, 1999). The idea is to bring together actors in terms of access to resources, literacy and power.

- **Partnership**

Partnership involves relationships between different players, governments, donors, community, and individuals. The term is used in many ways and is closely associated with participatory approaches. It is used to describe many forms of interaction between

diverse groups (Kelly & Vlaenderen, 1995). The synergies formed are important for the success of impact and sustainability.

- **Creating learning communities**

This has to do with a continuous learning process that has of late been embedded to enhance capacity to change and transform, that leads to transformation yielding changes in perceptions, thinking, behaviours, attitudes, values, beliefs, mental models, systems, strategies, policies and procedures (Balthasar & Rieder, 2000).

### **2.5.7 Overview and reflection of the evaluation literature**

The on-going discussion presents a situation whereby effective approaches of evaluation especially in LICs is methodologically complex with several challenges that hamper appropriate evaluation practice and application. Critical issues are still pertinent to the state of the art of evaluation, especially the LIC context. Notable issues arising were that:

- Technical sense of evaluation still persists to day, resulting in failure to accommodate value-pluralism, over commitment of the scientific methodology and a tendency towards managerialism. This disempowers managers have ultimate powers over evaluation process and utility of results, (Scriven, 1994).
- There is no one way to do evaluations as the knowledge of evaluation theory, research design, statistics and psychometrics is no longer sufficient as evaluators have to deal with complexities arising in LICs in terms of

management, policy, value economics and psychological questions (Joint Committee, 1981). Further standpoints are that:

- Each cross-cultural account makes sense in the particular context in which it was applied (Patton, 1985).
- Evaluations mainly serve donor interests and rarely do they attend to national interests, thus questioning the practical utility of impact evaluations in LICs (Asamoah, 1988).
- Results having distinct implications to policy are not always put to use (Asamoah, 1988).
- Evaluators are put in a context of completing values related to special interest groups, which are further and complicated in LICs which may reflect differences in epistemology and philosophy (Ginsberg, 1988).
- The LIC context is unique, providing unique challenges in terms of values to employ and methodological approaches to use (Child, 1977; Brass & Jolly, 1993).
- Criteria to use in determining the effectiveness of programmes in LICs remains unresolved (House, 1993).
- Guidelines on what values and criteria to choose from are still noticeably absent in the literature today to guide practice (Smith, 1991).
- Political expediency strongly influence decisions about policy and planning (Taut, 2000).

### **Theoretically:**

- Ideological disputes still prevail in terms of forms of evaluation c.f qualitative and quantitative paradigms, whereby evaluation has developed without an acceptable philosophy leading to diversity in evaluation theoretical perspectives (Guba and Lincoln, 1989).

- The literature has been accompanied by complexity that has proliferated growth of evaluation literature. This has been described as that which is ambiguous, with competing perspectives, and conflicting roles (Patton, 1982).
- That there is no univocal principles of evaluation any more than there are single, universally accepted philosophies of knowing or establishing the truth (epistemology) and how it affects practice; few theoretical process-modes have been developed, thus calling for additional research in the area (Scriven, 1994).

The view by Smith (1991) that differences should be seen as opportunities to enlarge evaluation worldview, and not as a set back is an acceptable invitation that, the field stands to gain greatly. Also, Africa provides an excellent laboratory for evaluation theory testing and refining evaluation research techniques. Smith further states that, maybe be, new forms of evaluation are emerging, thus providing a laboratory for cross-cultural theory and refining research on evaluation techniques.

Evaluation of social programmes of this nature specifically aspires to be an institution for democratising public decisions by making programmes more open to public scrutiny and deliberation through the process of dialogue and serving the interest of both practitioners and theorists. This puts the central focus on expanding theory development, institutionalising evaluation, creating genuine demand and incentives for evaluation, and providing infrastructure to support the move to legitimise evaluations. The challenge provided is great!



According to Mowbray et. al. (1998) evaluation has always had to deal with a number of constraints that have affected its practice. Arguments have been raised to the extent that if evaluation is to play its role in social development in the 21<sup>st</sup> century, the field may have to take a more *realistic* approach to carrying out its activities. Evaluation then is not an *end* but a *means* and a system that enhances generation of knowledge.

The framework set governs evaluation and must be placed within a contexts whose rationale is becoming that of interaction of programme strategies and policies and it's environment; reforms requiring good governance and performance tracking; and a tool for result-based management. Considering the context within which this frame is to be implemented, it further poses a genuine challenge to evaluation practitioners that transcends from operations across technical expertise, governments and international societies.

## CHAPTER THREE

### METHODOLOGY

The research sought to examine programme evaluation practice by carrying out a metaevaluation on nationally evaluated programmes related to education, specifically, curriculum programmes in Kenya. The evaluation context and process were examined using the *The Programme Evaluation Standard* developed by the *Joint Committee* of professionals as criteria for judgement. The researcher conducted interviews using an instrument developed based on the theory of evaluation literature. The process involved describing, analysing and critiquing the results and making recommendations that would prove helpful in developing appropriate evaluation policies and practices. Below is a framework of details on how the research was conceptualised, approached and carried out.

#### 3.1 Research Framework

A comprehensive framework is important in a study of this nature that involves content analysis of written documents. The following steps were taken into consideration in preparation for the research.

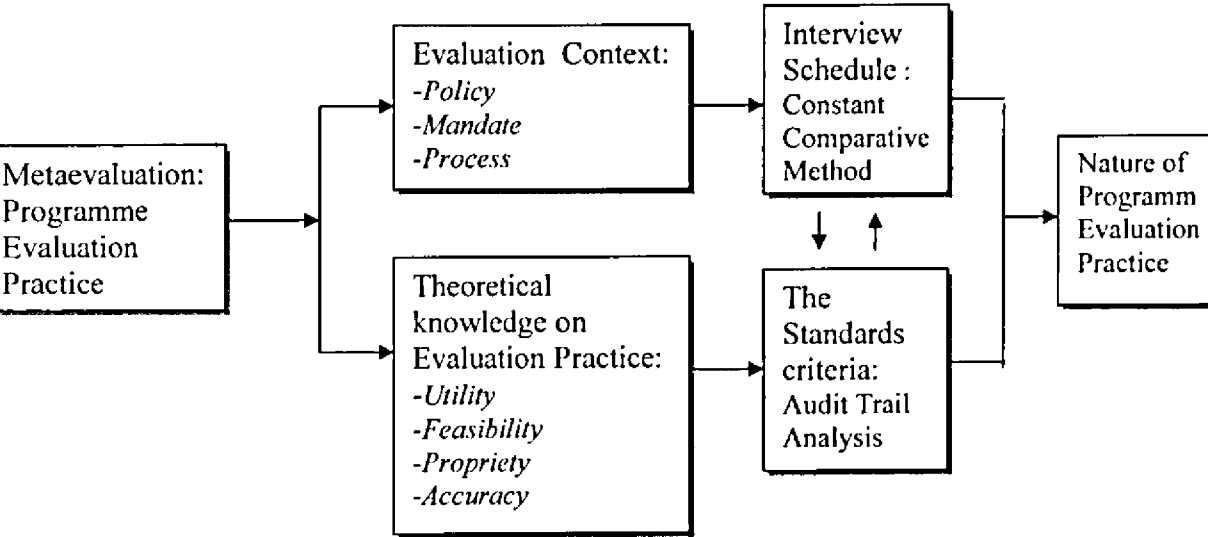
These were:

- i) Decision to use the Joint Committee of *Standards* (1994: from now on referred to as the “*standards*” or the Joint Committee\_JOS) as the criteria for judgements of worth of the theoretical knowledge of evaluation practice as a key instrument for the metaevaluation(Appendix: A)
- ii) Development of the *standards* analysis format guide to make the Joint Committee more applicable(Appendix: B).
- iii) Development of the interview schedule to determine evaluation context(Appendix: C)
- iv) Applying the *standards* criteria using audit trail analysis to determine the quality of evaluation practice based on theoretical evaluation knowledge and, carrying out interviews using the interview schedule to determine the context in which evaluation takes place and so evaluation context.
- v) Displaying the results from the *standards* analysis in a summary format by applying a judgmental scale given below.
- vi) Integrating the information from the *standards* analysis Criteria with the information from the interviews.
- vii) Critically analysing the results of the research and deriving recommendations.

**Table : 3.1.1 : Judgement Scale**

Attribute	Scale
Standard MET	<b>M:</b> When all conditions of the standards were met
Standard Partially Met	<b>PM:</b> When only half the key conditions were met
Standard Not Met	<b>NM:</b> When 75% of key conditions were not met
Standard Not Applicable	<b>NA:</b> Standard does not apply to the situation

**Figure : 3.1.2 : Illustration of Research Framework**



### 3.2 Feasibility of the Research Task

The question of *how, where, what* and *by what means* to examine the evaluation process was considered. Decision was made to carry out the research at the Curriculum Development Centre based at the Kenya Institute of Education (K.I.E). It is here that the data base from evaluation documents was derived. The Kenyan education system is centrally managed with a national curriculum, standard salary schedules for teachers across the country and a central Ministry of Education Science and Technology(M.O.E) that exercises authority by means of national examinations at the end of primary and secondary schooling. It has a number of legislated bodies one of which is the K.I.E.

The M.O.E has three important institutes charged with the handling of issues pertaining to curriculum and examinations. These are: the National Curriculum Development Center and Research(NCDCR), charged with developing the curriculum, improving content, publication of special materials for schools as well as conducting research and evaluations; The Kenya National Examinations Council(KNEC) which has the mandate to prepare, administer, mark examinations, issue certificates and conduct research and to evaluate students performance; and the Inspectorate which implements and monitors curriculum process. Evaluation is therefore a recognised process that is part and parcel of the education process (M.O.E, 1987).

K.I.E was selected because curriculum is the area in which there has been large national programmes commissioned for evaluation by the government. This brought in various

perspectives and players from scholars, parents, religious groups the civil society and the donors. Authority to conduct the research was requested and granted by the body responsible, which is, the Office of the President of the Kenya government(O.P).

Since formal evaluations generally culminate in written reports, these formed the basic source as the data base. Written reports of this nature generally reflect the processes and the nature of evaluations carried out which have been thought through by the evaluators. The advantage of this approach is that one can handle many evaluations conducted in a short period of time. The reports of this nature generally provide comprehensive information on the *accuracy* and *propriety* standards most of which could be obtained directly but less on *utility* and *feasibility* due to latent information. To make the results more valid, information beyond the manifest material on evaluation reports was supplemented through interviews to determine the latent information to complement secondary data arising.

The researcher acknowledges that the analysis resulting from documents of this nature and constructs from interviews is purely descriptive and judgmental in nature and that one cannot get away from the subjective nature of the research intended. It was important that validity and reliability measures be built into the research design to make the results credible. Throughout the research study, the researcher made and applied many reference points for judgement and well known ways of arriving at objective conclusions to ensure trustworthiness of the study. Measures of validity that deal with relevance, and reliability that is, consistency will be found in-built at various stages and levels of the research

procedures. This process is a procedure recommended by proponents of the naturalistic research design. The written reports are therefore useful and feasible way of investigating empirically the practice of evaluation.

### **3.3 Data Collection Instruments**

#### **3.3.1 The Joint Committee of Standards (1994) \_ Appendix: A**

Generally in such a study there would be need to determine the criteria for metaevaluation. A decision was made to use the Joint Committee. The task for the researcher was only to give information on the nature of the standards.

To determine the data base, the Joint Committee of Standards (1994; Appendix: A). was chosen for the analysis and applied following the Standard Analysis Guide (Appendix: B) developed by the researcher.

##### ***3.3.1.1 The nature of the standards***

The standards are organised around four important attributes of an evaluation namely, utility, feasibility, propriety and accuracy. The attributes are related to each other. For example, an evaluation which is not feasible, will not yield accurate conclusions; conclusions that are not accurate are not likely to be used; an evaluation which is conducted according to high *standards* of propriety will generally have much higher utility than the one with shortcomings in this respect. The attributes are a set of thirty (30)

standards each with an overview that provides definitions and rationale. The *standards* are grouped according to their potential contribution to each of the four attributes.

The Joint Committee (1994) states that the *standards* are not detailed technical standards and do not replace textbooks in research methodological areas such as data collection, neither are they equally applicable in all situations. They present *advice* on how to judge adequacy of evaluation activities but do not present specific criteria for such judgements. The developers further advise users of the *standards* that they (the *standards*) should not stifle the creativity of evaluators or impede the development of innovative approaches to evaluation; that they should be used as a guide for assessing evaluation plans and reports and not as a restrictive set of rules; that they should be used as a means of exchanging information about the quality of an evaluation between evaluators, their clients and other evaluators.

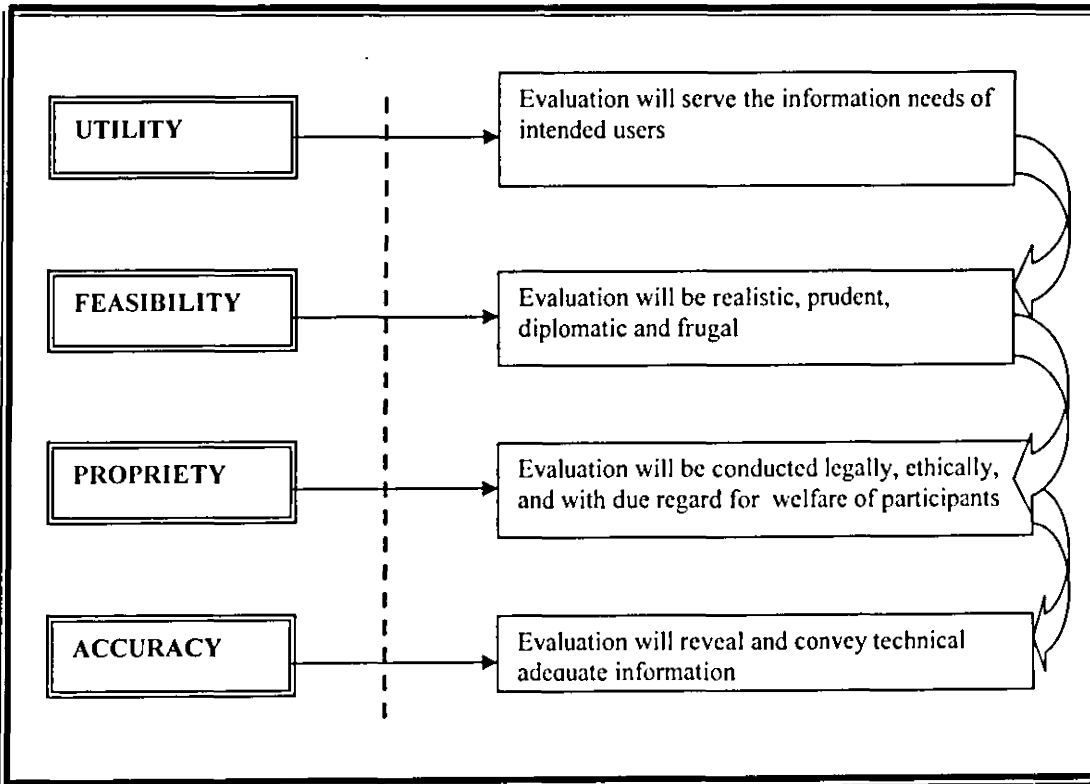
The researcher chose the *standards* as the criteria for determining quality in this study as they have become a milestone in the continuing developmental growth of programme evaluation as a discipline. Also, they do not align themselves to any particular approach. Since they are not geared towards any particular evaluation model or theory, they encompass a valid and widely shared conception of evaluation and conventional wisdom about evaluation practice in general, an attribute that appealed to the researcher. Also, the *standards* capture for the first time, in one self-contained source, the most critical principles of evaluation quality. No other *standards* for evaluation have been accredited



by the American National Standards Institute (ANSI), thus making the *standards* available world-wide (Hansen, 1994; Patton, 1994).

The researcher here shares the conception of the developers and in the same way perceives the *standards* as an appropriate criteria for the research, and that, despite its application beyond school settings, the *standards* were developed to deal directly with evaluation of education programmes which is the focus of this study. They are therefore an appropriate criterion for use in the study. The *standards*' attributes are detailed in the figure 3 below.

**Figure 3.3.1.1 : Joint Committee of Standards Criteria (JCS)**



Adopted from the Joint Committee of Standards (1994).

### **3.3.1.2 Developing a standard's analysis format \_ Appendix: B**

Written reports are useful and feasible way of investigating empirically the practice of evaluation. It was necessary to design the standards in a format that made it more applicable and practical for the task. A checklist of questions based on each of the 30-standards were developed resulting in the *standards* analysis format (Appendix: B). Information based on completed and written evaluation reports had it's content analysed in accordance with the *standards'* criteria.

### **3.3.2 Semi-structure face-to face interview schedule \_ Appendix: C**

In order to determine the evaluation context, a semi-structured face-to-face interview schedule was developed by the researcher based on the existing literature on programme evaluation (Appendix: C). The purpose was to help establish the mandate , or decree of evaluation; preparation measures taken when evaluation is called for; process used in conducting evaluation; and policy related to evaluation in Kenya.

The instrument was pilot tested and necessary changes made. This was done by doing a content validity check to determine coverage in relation to research questions for metaevaluation. The instrument was also sent out to evaluation authorities as well as to scholars to review and make comments on the extent to which the instrument have met the purpose set out for it, and that the questions as articulated were relevant.

The main advantage of the interviews was that it made it possible to capture aspects of evaluation practice lacking from direct analysis of the evaluation reports using the *standards* criteria.

### **3.4 The Research Design**

The research was based on the naturalistic inquiry paradigm, using qualitative data collection methods. The evaluation reports were examined using *content analysis* mode of research while applying the *audit trail* data analysis approach proposed by Guba and Lincoln (1981). The mode of research design prescribed is useful in socio-behavioural inquiry because of its interpretive nature that focuses on multiple realities, each representing a different perspective of the hermeneutical process.

#### **3.4.1 Content analysis**

Content analysis is a method of research used to analyse a wide range of communication that takes a variety of forms, one of which is the written form. The method is often used for producing descriptive information and cross validating research findings, a process that generally involves conceptualization, description and analysis of written material. The use of a framework to guide the analysis is an important step in the implementation of content analysis. Since the *standards* were already developed, the immediate task was to make the applications of the *standards* practical. This is exhibited in Appendix : B, Standards Analysis Format.

### **3.4.2 Hermeneutic dialectic procedure**

Where interviews were conducted, a hermeneutic dialectic procedure of interpretive and comparability of contrasting divergent views was employed with a view of achieving a higher level of analysis and synthesis through mutually shared exploration.

### **3.4.3 Trustworthiness of the research : Validity and Reliability**

Conditions of trustworthiness or validity and reliability considerations were ensured. Validity deals with the extent to which the measures represent what they should while reliability deals with consistency of measures.

The current study involves making judgements on written and human constructs material, using set criteria. Ambiguity in understanding what the evaluation authors meant or the criteria used for judgement were properly defined and understood, thus enhancing trustworthiness of the results. Various forms of validity in content analysis were identified. These were, *face, construct, content* and *predictive* validity. Face validity involves the correspondence between categories (and their units of context) and the concepts they represent as defined by the researcher through audit trail analysis. This was strengthened by carrying out interviews on the salient aspects of the reports and served as external validation. Rarely does content analysis deal with construct, that is, results correlating with other measures of the same, neither does it deal with predictive validity where results can be generalised to other contexts.

Distinction was also drawn between validity and specificity. Content that is specific, clear and unmistakable can be judged and categorised with ease and in a reliable manner. Content that is not clear (latent content) has to be implied from the text. Analysis can be done and what could be communicated in the reports for their manifest and latent content identified. Identification of manifest content may raise reliability, but, if it is the only content used, it may lower the validity of the findings. Identification of latent content may raise validity but they have lower claims of reliability. Thus to increase validity and reliability both contents were identified and analysed.

Three types of reliability are identified here. These are accuracy, stability and reproducibility. To be accurate, the judgmental descriptions of the evaluation practice as perceived in the evaluation reports must correspond to fixed *standards*; in order to be stable, the results from the analysis must be consistent over time; to be reproducible, those same results carried out by different researchers must be consistent, that is, inter-rater reliability. The researcher corroborated the results using an associate researcher trained on how to apply the *standards* criteria. The whole metaevaluation process was repeated or validated using random sampled reports from the same pool of the sixteen (16) reports (refer to sampling procedure \_ 3.5). This process was repeated after one month to establish the stability and so strengthening reliability of the results.

The process involving data reduction, data display, conclusions drawn and verifications were made explicit to show how the researcher ensured that the research was carried out effectively and efficiently.

## **3.5 Data Analysis**

Audit trail analysis procedure was used as well as the constant comparative method. The results were then integrated to determine the process through categorization and triangulation.

### **3.5.1 Audit trail analysis**

The researcher chose the audit trail as an appropriate standard of qualitative procedure, an acceptable process that ensures that an appropriate qualitative analysis is maintained. In this process, data analysis and interpretation follows closely on the heels of data collection. This is completed for one document and synthesised resulting in a more informed analysis before the next document is reviewed. This procedure was followed until all the documents for review were completed and a final synthesised overall review of all reports obtained.

The procedure that was used to synthesise the ensuing information the data was constant comparative method developed by Glaser & Strauss (1967). The process of constant comparative methods developed includes transformation and synthesis of all the data arising. The task is to transform the meaning unit given in the informant's everyday language and information from documents, into categories of statements using scholarly terms to describe the phenomenon.

The meaning units were then re-described and formulated to a general description of the structure underlying the variations in the meanings as it relates to the standards criteria. Data arising from different sources was also triangulated to enhance the interpretations.

### **3.6 Sampling Procedure**

Sampling in this case was not carried out for the purposes of generalizability or a representative population to which results were to be generalised. The sampling procedure was chosen to provide the broadest scope of information for achieving understanding of the phenomenon.

In the initial proposal, random sampling for selecting evaluation reports was made and a convenient (or purposeful) sample was proposed in selection of respondents. There was no problem in applying the convenient sample for selecting the respondents. However, there were problems in applying the random sampling strategy for selecting the curriculum evaluation reports presented by K.I.E. as the reports provided were sixteen (16). The total number of existing evaluation reports could not be established. However, not many more reports were expected as formal evaluation at the curriculum centre is not an everyday event and as determined from the research, formalised evaluations did not begin until the late 1980s. The researcher made the decision to review all the sixteen evaluation reports as the number in question could be handled in the period given for the study (Appendix F; Data source No: 1).

Also, the final results revealed that, the nature of the evaluation documents made them rather monotonous in the way they were approached and designed. This could be attributed to the typical research background of those charged with planning and conducting evaluations at K.I.E and the adherence to Tyler's objective oriented approach. Neither did reports have competing issues thus making it possible for the researcher to arrive at a consensus a situation not viewed as a limitation, as after all, establishing this was also the basis of the research.

Interviews involve human element of subjectivity. The respondents involved in the study had to meet set selection criteria that maximised trustworthiness of data. The criteria for selection were based on experience and knowledge of the investigative topic as well as one's association with the education process in general and if possible with the Ministry of Education (M.O.E) operations. A respondent to act as gate-keeper was identified. The gatekeeper was chosen on the basis of the person's elite and specialised position relating to his affiliation to curriculum centre and expertise in evaluation. The informants and associates from outside were proposed and identified through initial interview with the gate-keeper. However, the researcher determined the final list of respondents for the interview. The informants selected were also given an opportunity to further suggest other informants. The respondents at policy level included M.O.E personnel in administrative structure, the inspectorate, Teacher Advisory Centre (TAC) personnel, K.I.E personnel, university scholars, donor agents and education consultants.



### 3.6.1 Respondents for the study

Those interviewed are listed below:

**Table : 3.6.1.1 : Ministry of Education Officers**

Provinces	Nairobi	Nyanza	Coast	Total
Director Ed.				
PDE	1	1	1	3
DEOs	3	3	3	9
Head Teachers	4	4	4	12
Teachers	4	4	4	12
<b>Total : ( 36)</b>				

**Table : 3.6.1.2 : Kenya Institute of Education (K.I.E) Officials Interviewed**

Evaluation Research Officers	Gender		Academic Qualification	Extended Evaluation Training	Other Qualification
	M	F			
Officer 1	✓	-	Dr. In Ed.	✓	-
Officer 2	-	-	Dr. In Ed.	✓	-
Officer 3	✓	-	M.A Ed	✓	-
Officer 4	-	-	M.A Ed.	✓	-
Officer 5	-	-	M.A Ed	✓	-
Officer 6	✓	-	M.A Ed	✓	Data Analysis
Officer 7	✓	-	M.A Ed	✓	Data Analysis
Subject Panel and programme coordinator_ K.I.E 1. Primary 2. Secondary  Administrative _ K.I.E 1. Director 2. Academic Board					

**Table : 3.6.1.3 : Donor Organisations Interviewed**

	<b>Involvement</b>
World Bank	Project Implementation Unit (P.I.U)
British Council	Strengthening Primary Education (SPRED
JICA	Country Development projects
UNESCO	Science and Mathematics
CIDA	General
SIDA	General
NORAD	General
DANIDA	General

Those selected for the interviews cooperated with the researcher throughout the data collection, a condition desirable for naturalistic inquiry. The problem of availability came up as would be expected in any study especially personnel at higher levels of administration who generally have busy schedules. In order to keep to the study schedule, where it was taking too long to schedule an interview, an alternative officer who shared duties with the targeted officer and met the criteria for inclusion was recommended and designated to be a respondent. This was not common, however.

### **3.7 Data Analysis**

In this kind of study, data analysis and interpretation takes place while data collection is going on. The judgmental description of the documented material based on the four attributes of the standards was achieved through the process of audit trail using the constant comparative method of Glaser and Strauss already discussed until the last synthesised review of the last document was formulated.

The same process was followed for data obtained by interviewing respondents on salient aspects of the standards and information arising from the evaluation context. During the interviews, issues proposed by the first respondent regarding the standard were analysed by the researcher into an initial formulation using the constant comparative method.

Data analysis was completed before the second respondent was approached. While interviewing the next respondent, themes from the first respondent were also introduced where necessary in order to provide a critique. The central themes, concepts, ideas, values, concerns proposed in a previous interview from the previous analysis provided a comparison, thus leading to a consensus. This procedure was followed until the targeted respondents had been interviewed.

After completing data collection, half the documents were sampled for a second review after a period of three weeks to establish whether there was consistency in results. There were no contradicting issues arising and the summary of the results were in agreement with each other in terms of themes, values, and concerns on the standards criteria.

The two data sources, that is, information arising from evaluation report reviews and interviews formed the synthesised comprehensive judgmental descriptions to which assessments of strengths, weaknesses, overview, and common errors were applied based on the standards criteria. The resulting outcome as a result of the metaevaluation was then presented for judgement based on the scale of measure to provide a summary of the

outcome. The assessment related to each individual standard was transferred to a checklist as summarised analysis of the research results. The evaluation context was communicated in a written format.

### **3.8 Preliminary Preparation for the Research**

Data collection was done from September 1998 to May 1999. Certain operational procedures were put in place to initiate the research during the preparation to make the research process systematic in accordance with a scientific inquiry. Authority to carry out the research was requested and preliminary visits carried out at the research site, that is officers within the M.O.E administration and at K.I.E. The purpose was to negotiate cooperation that would ensure fully informed consent and building of trust which goes beyond methodological strategies but is key to success of hermeneutical, or dialectical process. All through the preliminary preparations and data collection process, there was willingness to cooperate. Being aware of incidences of unscrupulous facilitators and respondents who may well agree to cooperate because of being committed by their superiors but still retain mental reservations, the researcher instituted a stipulation requesting the respondent to publicly sign an agreement on their willingness to participate (Appendix: D).

The data collectors involved the researcher and associate researcher for the purpose of inter-rater reliability. The principle researcher provided, through training knowledge base and skills on the data collection process and analysis of data. The introduction of an

associate was to help in attaining credible results as this would provide an opportunity to corroborate the outcome of the research results.

The criteria for associate researcher involved being a holder of Masters degree with a thesis in education field. One other requirement was based on current involvement in research and or evaluation. The associate researcher was also trained in the application of the *standards* in order to arrive at a common usage. This involved the individual first reviewing the *standards* together with the principle researcher. The assisting researcher's performance was then assessed by asking them to apply the *standards* to an evaluation report already analysed to determine how well each *standard* had been categorized by the assistant researcher and to make notes on any difficulties experienced. The idea was to ascertain that the assistant researcher could carry out to acceptable levels, the research analysis as required.

# **CHAPTER FOUR**

## **RESULTS AND DISCUSSIONS**

This section presents the descriptive analysis of the outcome of the audit trail analysis and the specific determinants from interviews on evaluation context. The purpose of the research was to describe, analyse, critique the results and make recommendations that would prove helpful in the development of acceptable evaluation policies and practices.

### **4.1 Evaluation Context**

In this section, the descriptive analysis of the evaluation context derived from the interview schedules was analysed and presented. The evaluation context was based on the specific determinants given below. These were:

- Evaluation policy in Kenya
- Evaluation mandate
- Purposes for evaluation
- Objects evaluated
- Who conducts evaluation and in what capacity
- Who calls for evaluation

#### **4.1.1 Policies that govern evaluation in Kenya**

Various documents were analysed to determine the policy on programme evaluation. The results show that evaluation statements are found across central government documents and M.O.E documents. This shows that the usefulness of evaluation is recognised,

however, the analysis revealed weaknesses in the policy related to evaluation that could be referred to as incomplete. The central government itself states that despite the recognition and acknowledgement of monitoring and evaluation initiatives, there still exist major weaknesses in the twin management functions of monitoring and evaluation. The reasons given by the government were: lack of institutional setup for monitoring and evaluation, inadequate resource allocation, lack of stakeholder involvement, and capacity to carry out evaluation task (National Development Plan, 2002-2008).

A historical account of valuation reveals a situation where what purports to be evaluation is referred to as “research,” or left to an institution such as K.I.E to determine, yet it is an arm of the M.O.E and central government. As illustrated below, it is evident that evaluation was not yet conceptualised as it is approached in the evaluation literature. For example:

In The Ominde Report (1964)<sup>1</sup> .... Cap: 479, pg. 127, in the section titled *Research and Development*, there was recognition for an institution to determine worth of education programmes (this later became K.I.E) for wider facility for needed research due to inadequate existing knowledge on education projects....

In the Gachathi Report (1976)<sup>2</sup> .... recommendation no: 57, pg. 26, it was stated that there was need to support research activities in the country according to national development needs to provide necessary resources to carry out research activities....

In the Koech Report (2000)<sup>3</sup> .... K.I.E is considered but with no specific guidelines to evaluation process. At cap. 211, the statement made is general and it is with respect to any matters relating to the legal personality constitution

---

1. Ominde Commission. Education Report (1964). Kenya Education Commission report. Part I \_ Making evaluation relevant to post independent context.

2. Gachathi Report. Education Report (1976). Kenya Education Commission report, Part II \_ Establishment of the second university.

3. Koech Report (1999). Report of Commission of Inquiry into Education System of Kenya.

functions, powers and the general conduct and management of the institute as one may consider necessary or desirable....

The M.O.E has established bodies to support its functions for education and manpower development. There are eleven altogether. K.I.E, or the Curriculum Development Centre is one of the bodies. According to M.O.E, these bodies are semi-autonomous and have been established by Acts of Parliament with their own terms of reference. They each deal with specific professional areas. On analysing the main evaluation functions of K.I.E, evaluation did not feature as one of them (M.O.E, 1987). However, as one of the procedures for curriculum development, evaluation is listed as a function covering all aspects of education and the wider society in general. The departments directly involved in evaluation (and research within the curriculum area) were *Research and Evaluation Departmen* at the K.I.E and the subject panels. From the M.O.E central office, the Inspectorate, the other arm of the M.O.E also works together with K.I.E and it is responsible for monitoring of the curriculum implementation and monitoring as designed for schools.

One therefore sees a system with no clear guidelines on evaluation or legislation pertaining to evaluation per se. There is knowledge of evaluation function and acknowledgement of its importance but that there was reluctance by the central government (not necessarily the officers in charge of evaluation, such as K.I.E) to legislate evaluation and make it part of the mainstream government function. Things seem to be changing, however, as demand by the public and donor community on



accountability increases due to the insistence on *results based management* (RBM) and *transparency* which has become a global call in both LICs and ICs.

With the launching of the National Development Plan (NDP: 2002-2008), the recognition of the need to strengthen the monitoring and evaluation component as evaluation is generally referred to in the documents, has increased due to the need for accountability built into mainstream budget process and policy. The seriousness with which this is viewed is noted by the intention to set up a National Monitoring and Evaluation Committee (NMEC) made up of all Permanent Secretaries within the government ministries, Director of Planning/ Economic Secretary, central planning units of the line ministries, Director of Budgetary Supplies Department and Budgetary Monitoring Department (BMD). It is the intention of the central government to establish the Central Planning Unit (CPU) which will be established to serve as a secretariat to the process. NMEC will have the *mandate* to ensure compliance. In education the government recognises that the curriculum relevance has been *ad hoc* while monitoring and evaluation of programmes have been absent. Evaluation has thus been proposed as one of the priority areas for curriculum programme development (Government of Kenya, 2002-2008). Even then there was no indication in the document that these intentions will be legislated as an act of parliament.

In later discussions it becomes clear that donors and activist civil society groupings do have great influence on compelling the process of evaluation in the country. The interviews show that donors can compel the government to carry out an evaluation

(whether it is appropriate or not) by tying evaluation to aid. The donors' mandate on evaluation and social programme improvement could play a role in influencing evaluation policy but this depends partly on the donors priority at the time. It is also not always that donors' adhere to evaluation principles. The lack of systematic approaches to evaluation by donors results in fragmented evaluation intervention. This may imply that the donors do not make reliable impact on evaluation policy nor do they aid establishment of a culture that embraces evaluation demand that is technically appropriate, definite, long-term and appreciated.

One could therefore summarise the policy issue on evaluation as incomplete. The evaluation functions are not translated into programme activities or operations, nor is evaluation appreciated as useful and far-reaching tool in areas of programme development and human endeavour. Evaluation has yet to be institutionalised in the public sector in accordance with acceptable state of the art of evaluation literature. The government seems to look at evaluation as *criticism* and not as a tool for social development and reform thus leading to repercussions. Further still, those who create situations for evaluation in themselves have not applied evaluation principles that are sustainable. If anything, the demand for evaluation by the created synergies could be said to be driven by self-interests as well as issues beyond their control and not necessarily evaluation principles, methods and approaches.

#### 4.1.2 Establishing the mandate for evaluation

To establish the mandate, the researcher interviewed various groups and persons as reviewed existing information from the media and government documents. The results showed that various factors do contribute towards mandating evaluation. There was pressure by the stakeholders (not the government) to carry out evaluation and implementation of the results.

The stakeholders were:

- politicians from various political parties
- civil society, such as religious groups\_ mainly the Anglican and Catholic churches
- non-governmental organisations involved in education
- research institutions, such as Institute of Policy Analysis and Research (IPAR)
- donors such as the British Council and the World Bank

Since Kenya gained independence from Great Britain in 1963, there has been a number of commissioned reports on education. The Ominde Report (1964) sought to make a difference between the colonial (pre-independence) curriculum and the post independence curriculum. The purpose was to make the curriculum programme *relevant* to the local needs. The interviews revealed that in both cases, commissions were set up to review the curriculum but no formal evaluations were carried out.

In 1981 the government set up yet another commission, The Mackay Commission to advise on the establishment of a second university (besides the University of Nairobi). Apart from a new public university, now known as Moi University, the Mackay Commission also recommended the current model of education system of 8-years primary, 4-years secondary and 4-years university (8-4-4). The system was thus charged

from 7-years primary, 4-years secondary, 2-years A-level and 3-years university learning(7-4-2-3). The 8-4-4 education system was implemented in 1986. The interviews revealed that formal evaluation of the curriculum began after the implementation of the Mackay Commission recommendations in 1985. Since then there has been, indeed, a national approach to evaluation in Kenya.

After the Mackay Report (1981), there was the Koech Report (2000) which looked into the country's education needs at the dawn of the 21<sup>st</sup> century. It was during the Koech Report that evaluation was finally addressed as it is purported in terms of its professional function as an essential guide to decision making. As a result, a needs assessment was conducted by the K.I.E to have an input into the commissions work so as to enable it address the country's curriculum needs.

The government has been drawn into evaluations after claims that the 8-4-4 system was too demanding in terms of the work load for both teachers and pupils. It has also been argued that these inadequacies of the 8-4-4 system eventually lead to high education costs and expenses to the parents as a result of the cost sharing policy. These criticisms have been made by stakeholders since 1989 when the first formal evaluation was carried out but the government did not act to implement the results(Daily Nation, 27<sup>th</sup>. March 2000). Lack of implementation of evaluation results is reflected in various evaluation documents that were reviewed. The government of the day was reluctant to evaluate a system it set up and implemented. Statements coming from the government angered those advocating for change more so after an evaluation had been carried out. The said Daily Nation

carried a story titled, *8-4-4 will stay, says Minister of Education*, in response to the recommendation to overhaul the 8-4-4 system with a view of reverting back to the 7-4-2-3 system of education. Below is a sample of the comment of the Minister of Education while officiating at a public function:

*".....the recommendation would not be implemented.... that since the committee was appointed by the government, the government had the choice to implement or reject its recommendations..... that Kenyans were not in a crisis to rush into implementation since the ministry was amending the education act.....that 8-4-4 system was here to stay and the government will not scrap it.....that those who were opposed to the 8-4-4 system of education are critics of the president Daniel Arap Moi (president of the Kenya government: 1978-2002) since he implemented the 8-4-4 system....."*

(Daily Nation, 27<sup>th</sup> March, 2000)

This mode of attitude by the government resulted in interested groups such as donors creating conditions that would tie reviews or implementation of the evaluation process to aid or donor funding. The interviews revealed that there were threats of withdrawal of education funds unless the outcome of the evaluation was implemented. Britain's Department for International Development (DFID) had given the government up to the end of the year 2002 to effect curriculum changes arising from evaluations or else it would withdraw grant to education projects. This was also cited in the other national daily newspaper, the East African Standard (13<sup>th</sup>. April, 2002). The interviews revealed that, as a result of the government's insistence on their way of doing things despite public outcry, the donor community tied aid to issues such as programme review, threatening that aid to the M.O.E or the government in general, would be discontinued unless evaluation was carried out and results utilised as desired by the public. The interviews revealed that co-financing curriculum was not only threatened by government's reluctance to implement evaluation results, but also by gross violations of human rights issues. The Kenya Human

Rights Commission criticised the government and was supported by international bodies, such as the World Bank who maintained that any future aid would depend on both political and economic accountability on the part of the government.

The researcher, however, views the donors' approach as non-systematic nor principled towards the function of evaluation. Tying evaluation to aid is not desirable for appropriate evaluation practice that guides professionalism because the donor is more focused on conflict between the public and the government and not the function of evaluation.. Also, bringing issues that are not directly related to the evaluation need, that is, programme development, is in itself a conflict of interest. Consider accusation by the World Bank on the country's human rights issues which may or may not have been relevant to programme development needs at the time.

In view of economic difficulties being experienced by both ICs and LICs, Kenya included, the insistence on restructuring 8-4-4 system of education could also appear far-fetched and unrealistic because of Kenya's financial constraints. This may mean that the stakeholders too, do not always make realistic and informed decisions on evaluation issues, nor do they realise that results do not have to be put to use as they arise but can be incremental. Despite pressures from various interest groups, the stakeholders still had to wait for the government to give a go-ahead and approval to carry out evaluations, and, implement the evaluation results. The above scenario reflects an evaluation mandate that is generally created more by political and technical needs other than evaluation methodological needs, such as, monitoring and formative and summative evaluations. It

seems that evaluation is not mandated for its function, such as, programme development for improvement in terms of its monitoring, formative and summative function.

The scenario, reveals a situation generally controlled by politics other than by professionalism. Having a centrally controlled system of governance, those charged with evaluation have had to abide by the regulations laid down by the central governing body of the Kenya government. The system depicts a hierarchical type of administration with rigid lines of control (maybe not clear), one way flow of information and bureaucratised organisation whereby the evaluator has little say over the evaluations. This kind of setup reflects a managerial position that characterised the first generation of evaluators where the manager decides when to evaluate, what to evaluate, when to carry out the evaluation and whether to use the results. This kind of practice leads to such undesirable consequences that the evaluator cannot be held responsible for failure of evaluations and the manager has ultimate powers. Such practice generally leads to disempowerment. The Kenya evaluation system presents such a situation where the government is the body that determines:

- the kind of evaluations to conduct: summative, formative, diagnostic, internal or externally conducted.
- when in the life of the programme an evaluation should be conducted.
- the finances involved in facilitating the process of conducting evaluations.
- the dissemination and utilisation of evaluation results.

The evaluations arising from evaluations carried out over time were needs assessment, formative and summative evaluations.

### 4.1.3 Purposes of evaluation

In determining the purpose of evaluation, two aspects came up, that of socio-political demands and programme development needs.

In terms of socio-political demands, it was established that power and accountability relationship between the government and various bodies such as political opposition bodies, donors, the general public and scholars were key to what drove evaluation to be undertaken. The aim of evaluation in this sense was subject to a diverse range of influence and interest beyond an evaluand's theoretic base. Evaluation was also looked at as a means of solving socio-economic needs and ills.

As a result, evaluation ends up being sanctioned not for its technical function but to meet demands instigated by various conflicts that arise whenever programme evaluation is called for nationally. Therefore evaluation gets sanctioned by donors should the government not comply. On the other hand, the government opposes evaluations to defend the programme it has set up. Consequently, the following are the purposes of evaluation in the social-political context:

- i) To meet pressure demands of stakeholders by various interest groups.
- ii) Putting government to task over its programmes and policies.
- iii) A condition for continued donor support to the country.
- iv) The government protects its stand on programmes it has set up despite problems with impact of these programmes.
- v) To engineer policy change (e.g reverting to 7-4-2-3 from 8-4-4).
- vi) To help solve social needs, for example, making education affordable.



The second element that defined the purpose of evaluation process was programme development needs. These were:

- i) Need for *relevance* and *content* of programme arising mainly from stakeholder concerns (not the client/ government).
- ii) *Reviewing a programme* for restructuring.
- iii) To justify need for *policy change*.

The outcome of the audit trail analysis, that is, analysis of the evaluated documents using the *standards* resulted in the following specifics in terms of purposes for evaluation.

These were:

- i) Development and improvement as it related to the evaluand.
- ii) Creating awareness and knowledge.
- iii) Aiding decision making.
- iv) Issues on the evaluand process and procedures.
- v) General accountability of programme intents
- vi) Cost effectiveness.

(Table : 4.2.1.1 ; Appendix F : Data Source No. 2)

Details on the arguments that arise in trying to carry out an evaluation and to designate its purpose was found in various sections of the report. Some of the issues that came up were that, there was no involvement by stakeholders in deciding what to evaluate; the purposes were not well thought out and so were not comprehensive enough in terms of what led to developing the evaluand and making its developmental process efficient; the extent to which the purposes were formatively or summatively stated was raised. It was established that, the statements of intent did not give direction towards defining the intended activity.

#### 4.1.4 Objects evaluated

In determining the objects for evaluation, the social context as well as the theoretical base of the evaluand were analysed.

In respect to theory base of the programme of inquiry, that is, technical values underlying the object being evaluated, analysis took into consideration generally accepted theory base of the evaluand. Key elements of the evaluand based on theory of the foundation were covered as ascertained by McCormic & James (1990). This was the curriculum *intended*, in terms of *what*: education objectives, subject area objectives and content; curriculum *taught*, in terms of, *who*, *when*, *where*, *how* and by *what* means; and curriculum *attained*: or achievement.

However, the social aspects of the evaluand which is important as it defines the context in which the evaluand exists, was not addressed. The M.O.E spells out a system of education in terms of its goals that ought to remove shortcomings related to social aspects of the evaluand or the education system, that is, how relevant the programme is; a system that offers equal opportunity for all; a system that instills in its citizens a sense of nationhood; a system that shares common problems and aspirations of the international community (M.O.E, 1987). These objects of concern were not evident in the results arising from the metaevaluation analysis.

Specific objects of evaluation as it relates to curriculum included the following:

- Education/ school philosophy
- Curriculum programme
- Implementation
- Subject area \_ course units taught
- Personnel\_Teachers
- Reactions to programme process

(Table : 4.2.1.2; Derived from Appendix F, Data Source No. 2a)

The results show that the objects for evaluation can be tailored to address the theoretically conceptual areas of the evaluand. However, the required information did not address the effectiveness; harmful side effects, costs, responses to learner needs, and meaningfulness of assumptions made, issues that are generally desired within the *standards* criteria. The questions addressed did not appear to have been prioritised either.

It was also established that the emerging information arising was important to significant stakeholders having a direct affiliation to the evaluand. These were teachers, learners, parents, board of governors, and the community. It was not possible to establish the extent to which the information sought responded to the needs of the stakeholders as their views were not formally sought much as the media carried public opinion on some of the pertinent issues. However, in comparing internal and external evaluations, there were more pertinent objects addressed in external and needs assessment evaluations in terms of the education system and education historical development over time.

The client of evaluation is the central government through the governing bodies, M.O.E and K.I.E. Both governing organs generally make decisions on object of evaluation and nature of evaluation; K.I.E advises the M.O.E on approaches to take.

Despite the cited inadequacies, the scope of evaluations appear to address acceptable evaluation questions needed to guide programme development and reform to a large extent. However, more thought needs to be put into deriving information scope and development of evaluation knowledge from one evaluation to the next. The evaluand also needs to be considered within the social context in which it exists.

#### **4.1.5 Who conducts the evaluations?**

This question looked at in terms of academic background, experience with subject matter and evaluation methodological knowledge.

The interviews reveal that all the staff of K.I.E who were responsible for internal evaluations had basic research background and at least a Masters degree with two officers holding doctoral degrees in education. All had training in at least one course in programme evaluation.

The external evaluators had even higher professional credibility as some were professors who had conducted research in the area for five to ten years. This reflects competence as exhibited in subject area training, knowledge and experience in research.

The ensuing results show that the team had substantive credibility to defend utility, practicality, integrity and technical adequacy in terms of social science research principles. However, results reveal that there were problems with evaluation methodological approaches as reflected in a number of *standards*, for example, U1\_ *Stakeholder Identification* where information needs were not sought; U7\_ *Evaluation Impact* in terms of the nature of recommendations and quality of the statements made. This deals with issues of whether the statements were speculative and specific or formatively or summatively oriented. None of these criteria were met. Problems of not setting out conditions that would accommodate *propriety standards* such as P5\_ *Complete and Fair Assessment*, and P6\_ *Disclosure of Findings* also came up, in that, the standards related to the protection of the rights of individuals were not met.

This led to the conclusion that, those charged with carrying out evaluations were not competent in the area of programme evaluation methodologies which in an evaluation process, goes beyond a technically social science research process and procedures. From the foregoing discussion, it would appear that academics, relevance of professional area and research experience, play a role in determining who carries out the evaluation, but not expertise in evaluation methodology.

**Table: 4.1.5\_1 : Typical Composition of Evaluators**

Internal Evaluators	Summative Evaluators
Research and evaluation officers (4) Programme coordinators (Pri/Sec) Subject specialists(pri/Sec) The Inspectorate	University scholars Each evaluation had 4 evaluators

#### **4.1.6 Who calls for evaluation and funds evaluation?**

For the actual evaluation to take place, it is the government's prerogative to order the evaluation despite the pressure from various interest groups. The analysis carried out so far reflects a situation where the government, was first and foremost sanctioned to carry out evaluation, or influenced through pressure from various interest groups. One could therefore conclude that though the government does not actually influence evaluation mandate, it is the government, through the central governing body who makes the final decision for external evaluations to be carried out. However, if the decisions to be made do not affect the entire whole parts of the programme, then internal evaluations could be carried out without seeking permission from the central governing body.

As regards the *funding*, the government has set up a fund purposely for evaluation purposes for the Kenya Institute of Education (K.I.E) to carry out its day-to-day functions of evaluation. In the event that a large scale evaluation is to be carried out, both the government and other organisations, such as donors, generally collaborate to fund the exercise.

#### 4.1.7 Determining preparation for evaluations

In preparing for evaluation, it was important to understand the line of authority and governance, resource mobilisation, evaluation methods in terms of data collection approaches, terms of reference, preparation of reports and their annexes, as well as conditions to ensure optimum utilisation of the results. Since all other attributes are discussed in detail in the sections that follow, this section will only highlight some of the issues and will concentrate on the line of authority and governance of evaluations.

In Africa, government systems are generally characterised by hierarchical systems of governance where the head of state has the final word in any policy matters

(Traut, 2000). The preparation of the programme was viewed in terms of distribution of roles, responsibilities, allocation of resources through interviews and reviews of M.O.E documents. The review documents revealed that the *Permanent Secretary* in M.O.E is supposed to be in charge of all government policy matters in M.O.E, including evaluation in all ministries bodies. Reference was made to K.I.E being a semi-autonomous body, however, K.I.E also falls under the division of Education Policy and Programmes (EPP) which according to the documents is charged with formulating the policy for K.I.E. and 8-4-4 system of education. One therefore sees a possible mix-up in the line of authority whereby on one hand, K.I.E is recognised as an institution through an act of parliament and at the same time, it is under the umbrella of other sections of the ministry from which is it expected to take instructions. At any rate, there is yet a higher authority, the central government, which is headed by the President of the Republic of Kenya who is in charge of governance of all government ministries. This shows that K.I.E may not be able to

make independent decisions to implement outcomes from national evaluations involving professional evaluation needs.

The results show that the source of information at K.I.E varies, ranging from government education publications, newspapers, and subject personnels. The secrecy with which government documents have been guarded makes access to most of the evaluation documents difficult for the general public except for government documents sold by the Government Printer, such as the education reports. However, the body does not publish evaluation reports. Even with the authority by the research clearance house based in the Office of the President that authorises researches in Kenya, government officers (except those at higher administrative levels) were not sure whether they could issue evaluation documents even when the researcher was authorised. In fact, a situation that arose resulted in a situation where even the education officers within the same M.O.E did not have knowledge of or access to the evaluation documents. A statement such as:

*.... " I have never seen a K.I.E or government evaluation report..... "*

was common during the interviews. Out of twenty-four (24) respondents interviewed, twenty (20) answered in the negative. This presents a situation whereby not all government documents, especially evaluation reports are circulated. As will be seen in the discussions later, U *Utility Standards* that deal with dissemination of evaluation results were not adequately met. This raises the question of the appropriateness of the dissemination process, an important component of the evaluation process as it helps in facilitating distribution of results to the stakeholders for eventual utilisation. There was



also the implication that information could not be collected, according to the terms of reference, even with authority. Fear of repercussions when those in authority felt offended or intimidated when an evaluator was interacting with respondents during information collection was a salient factor. This could not be easily ascertained from the interviews as the interviewees, such as K.I.E personnel, declined to answer the question.

#### **4.1.8 Research procedures**

The analysis shows that the evaluation procedures to be followed when conducting evaluation were not in-built into government programmes. The interviews revealed that those responsible for conducting evaluations, such as K.I.E, were however knowledgeable about the various modes of evaluation and the professional timing for them. However it was not possible to apply professionalism when deciding to carry out evaluation or when one saw the need to conduct an evaluation. This was because of the political expediency displayed in the Kenyan governing context, more so, when carrying out evaluations of large programmes such as 8-4-4 system of education. On interviewing the K.I.E evaluation (and research) personnel, the situation was such that, whereas it is the central governing body (verses the evaluator) who mandated formal national evaluations, whether formative, summative or needs assessment, it was only after pressure from civil groups and donors.

The internal evaluations were not utilised to change policy but to influence policy change over time. The K.I.E staff often worked together with the schools and the Kenya National Examinations Council (KNEC) through their research department, one of the legislated

departments of the M.O.E that acts on day-to-day needs of curriculum review and development. These internal evaluations were used to strengthen the curriculum process internally and without publicity. Interviews revealed that policy decisions on programme evaluation were generally effected much more when the central government commissioned an *external evaluation* which also attracted a lot of publicity from interested parties. Even then the evaluations were *not timely* to provide desired programme development or to influence programme change.

## **4.2 Theoretical Basis of Programme Evaluation Practice : (Analysis based on the *Standards*' Criteria)**

In this section, the descriptive analysis and outcome of the audit trail as it pertains to the *standards* criteria will be analysed and elaborated. A decision will be made to determine the extent to which the *standard* was met. A description of each the *standard* will be given at the beginning of each attribute<sup>4</sup>.

### **4.2.1 Descriptive analysis of the *Utility Standards***

Utility is often mentioned as the most important quality of an evaluation procedure (Forss & Carlsson, 1997). The utility standards define whether an evaluation serves the practical information needs of intended users. They guide evaluations to be *informative, timely and influential*.

---

<sup>44</sup> The descriptions of what each *standard* entails as described at the beginning of each *standard* is derived directly from the descriptive statements as they appear in the *standards* publication. The acknowledgement will only be noted and referenced when introducing the *standard* attribute.

## **U1 \_ Stakeholder Identification**

The *U1\_standard* requires involvement of multiple and diverse stakeholders in planning of evaluations, data collection, as respondents, those influenced and affected by the evaluation results. A review of the documents revealed involvement to stakeholders during data collection depending on whether the evaluation was driven by internal or external demand. Generally, students, teachers, parents, board of governors and the community within the vicinity of the school were considered as respondents intended and targeted stakeholders; categories and units of analysis (Tables: 4.2.1.11; 4.2.3.1. respectively).

The external evaluation reflected a larger scope of respondents that included school-leavers, school dropouts, civil activists, such as the religious groups, students with special needs, departments of M.O.E, such as K.N.E.C, the national examination body and T.S.C or teachers affiliate body. The results indicate that those at high administration policy making level were not contacted or interviewed. Those affected and influenced by the evaluations were identified by analysing the statements for recommendations made and included mainly policy makers at the M.O.E, programmers at K.I.E., teachers, and schools. The special groups category was examined by looking into units of analysis and cross-cutting issues for more heterogeneity.

The contents of the report could therefore be viewed as narrow, as it relates to internally sanctioned evaluations compared to external evaluations. This is a problem as it could lead to limited use. Contacting those with vested interests who have direct influence on policy is important as this influences the *use* of evaluation results leading to a more responsive evaluation.

Key areas were presented thereby widening the scope of stakeholder perspective. These included regional divisions, such as those in the urban and rural, school types, class levels, and course units that are important in broadening the scope of stakeholders. However, the interviews did not reveal any indication that the evaluation information needs of stakeholders were sought *while planning* for the evaluation, neither were the stakeholders involved in data collection. This may mean that the views of respondents in this respect, as well as those affected and influenced by evaluations were not taken into consideration. Interviews revealed that the information needs of the evaluation were derived by those who commissioned evaluation, that is, the government and K.I.E. Those who co-funded evaluations also had a say, for example, donor agencies, such as the World Bank group and DFID.

Since the stakeholders were not approached to give their information needs while evaluations were being carried out, it was difficult to determine whether their needs arising from the evaluations were actually met. The researcher considered the information needs of stakeholders from the views often expressed in the media. One of these was the need to review the 8-4-4 curriculum by reducing the curriculum content as

well as the number of subjects taken in order to decrease the workload for both teachers and students, a problematic area frequently occurring in evaluation outcomes. Much as this particular issue was consistent with the audience needs as described and handled through the design of the evaluation, the evaluators did not deliberately solicit stakeholders' evaluation needs.

### **U1\_Standard was only Partially Met (PM)**

#### **U2\_ Evaluator Credibility**

The Joint Committee of Standards recommended a team for conducting the evaluation since few individuals possessed all of the characteristics needed to accomplish an evaluation task.

The issue as to who conducts evaluations has been addressed. It was concluded that, credibility of the those carrying out evaluations was met as far as academic qualifications experience and social science technical competence were concerned. However, expertise in terms of evaluation methodology was not met.

Other issues that needed to be addressed in U3\_Standard as desired in the analysis guideline included the need to involve the stakeholders. The results reveal that, the audience of evaluation addressed in U1\_Stakeholder Identification were generally not informed about the progress of the evaluation, neither was the audience given a chance to react to the work plans and composition of the evaluation team. Credibility was

determined entirely by those commissioning the evaluation. Glaring short comings were identified all across the Joint Committee. This had implications in meeting the U2\_ *standard* whose attainment relied on attaining other standards, for example, U1\_ *Stakeholder Identification*, if the needs of the persons affected or influencing the evaluation are not taken into account; F2\_ *Political viability*, if the evaluation is not conducted in anticipation of positions of key interest groups that influence policy; A2\_ *Context Analysis*, if the context is not examined in detail; U3\_ *Information Scope and Selection*, if pertinent questions of the programme being evaluated as designed and selected are not asked. None of these *standards* were met.

## **U2 \_ *Standard* was Partially Met (PM)**

### **U3 \_ *Information Scope and Selection***

The standards recommend that information should be broadly selected and should be relevant to decision makers' objectives, significant to stakeholders and sufficiently comprehensive to support a judgement worthy of merit.

In earlier discussions an analysis was carried out to determine the purposes for which evaluations are carried out (and the objects evaluated). The U3\_ *Standard* addresses these issues but will only look into issues which the *standard* did not previously handle. The conclusions reached in earlier discussions reflected the following arguments, that,

- Issues generally addressed looked into development and improvement of the evaluand as well as creation of awareness and knowledge.

- Evaluations may not have been seriously thought before selecting the scope and issues to address.
- The theoretical base of the evaluand was adequately identified and addressed as regards the key elements.
- The social aspects of the evaluand were, however, not addressed.

The magnitude of the most commonly occurring purposes of evaluation are given below:

**Table : 4.2.1\_1 : Commonly occurring purposes of evaluation**

(Derived from Appendix F, Data Source No. 2a)

<b>Purpose</b>	<b>F</b>	<b>S</b>	<b>N</b>	<b>Total %</b>
Development and improvement	19.4%	3.2%	0%	<b>22.6</b>
Awareness and knowledge	32.2%	16.2%	23%	<b>71.0</b>
Aiding Decision Making				<b>Other 6.4</b>
Planning curriculum process				
Accountability				
Cost Effectiveness				
				<b>100%</b>

**Key: Formative (F) Summative (S) Needs Assessment (NA)**

The extent to which the objects were addressed were as follows:

**Table: 4.2.1\_2 : Objects of evaluation**

(Derived from Appendix F, Data source No. 2 a)

<b>Objects</b>	<b>Total(%)</b>
Education/ school philosophy	<b>38.7</b>
Curriculum programme	<b>29.0</b>
Implementation	<b>12.8</b>
Subject area _ course units taught	<b>6.5</b>
Personnel _ teachers	<b>6.5</b>
Reactions	<b>6.5</b>
<b>TOTAL : 31</b>	
	<b>100%</b>

What was also common across the evaluations was the tendency of commonality on the issues addressed to the extent that, despite the differing needs of evaluations, they almost always related to the following programme needs or in this case, curriculum needs:

- objectivity and relevance of the programme, that is, school curriculum
- issue of implementation or process
- available resources and facilities in terms of effectiveness and efficiency
- teacher quality as it relates to competency of direct implementers, that is, the teacher

Rarely did evaluations address the following issues related and desirable for programme needs:

- analysing subject content and sequence
- reasons for regional disparity, such as urban/rural differences
- equity studies across programme achievement and process, such as gender relations
- Complex processes of the programme such as classroom learning processes and classroom interaction
- emerging national issues of programme needs and process, such as influence on school dropout and costs



- Key programme measurement descriptor, such as student performance indicator, that is, test scores

Furthermore, evaluations do not address the following:

- subject choice and combination leading to student career development
- cost effectiveness

Other areas that could sound far-fetched but are part of the curriculum process and are related to the social needs and challenges, such as those identified in the Kocch report, were not addressed. These include HIV /AIDs as well as alternative and continuing education, management and coordination of education.

The researcher is therefore of the opinion that the evaluations may not be seriously thought through before considering the scope and selection of issues to address. Much as the issues addressed are pertinent questions of the programme and the evaluand, there are notable omissions that may hamper maximum benefits for those affected and influenced by the programme and outcome of the evaluations.

The quality of evaluation, as approached, was also determined in terms of the way in which the evaluations were stated as formative, reflecting immaturity, or summative, that is, reflecting attainment. Table 4.2.1\_3 reflects the results of the analysis.

**Table : 4.2.1\_3 : Quality and characteristic of statements of purpose**  
(Derived from Appendix F, Data source No. 2b)

<b>Evaluation Function</b>	<b>%</b>
Formatively stated:	68.8
Summatively stated:	66.7

Formatively stated : maturing  
Summatively stated : attained

The quality of evaluation was analysed by reviewing the objectives as stated to determine whether the evaluations were given direction when stating the purpose. The results show that more than one third of the time, the evaluations, as approached, did not have the purposes evaluatively stated.

There was no clear indication of the review of pertinent literature and a comprehensive review of previously conducted evaluations to provide full background account and continuity from one evaluation to another in terms of programme process and development. However, in comparing internal and external evaluations, more pertinent issues were addressed in external evaluations and the needs assessment in the education system and its historical development over time.

The literature on external evaluations generally presented the following:

- historical account from traditional to introduction of formal education systems
- the curriculum content then \_ agriculture and industry for Africans and academics for Asians

- pre-independence education review committees as guided by various education reports:
  - Phelstopes (1936); Pin (1936; Beecher Report (1949); Binns (1952).
- post-independence\_ (after 1963):
  - Kenya Education Commission or the Ominde Commission (1964)
  - The National Committee on Education and Objectives and Policies or the Gachathi Report (1975)
  - Presidential Working Party on second university, the Mackay Commission (1981)
  - Presidential Working Party, the Koech Report (2000).
- the listing of government institutions concerned with the curriculum process

Depending on whether the evaluation was on secondary or primary level of education, the goals of the education level were presented and the subjects taught, listed. This was then followed by purpose and objectives desired by the evaluation. Whenever previous evaluations were addressed, the typical problems within the programme such as a loaded curriculum, was stated but without comprehensive discussion of previously conducted evaluation results that could guide the reader. Furthermore, theoretical concerns regarding education and the curriculum evaluation process and development were not addressed nor did the reviews look into issues around programme development and process generally, or in Kenya in particular. There was no reference to curriculum evaluation researches, discussions with stakeholders to provide background knowledge or anticipated information that emerges when addressing evaluations, or the curriculum reviews.

Despite these inadequacies, the scope of evaluations appear to have addressed acceptable evaluation questions needed to guide programme development and reform. However,

more thought needs to be put into deriving information scope and development of evaluation knowledge from one evaluation to the next.

### **U3 \_ *Standard* was Partially Met(PM)**

### **U4 \_ Values Identification**

Value is the root term in evaluation and it attempts to rate an object on its *usefulness*, *importance*, or *general worth* and therefore, the fundamental task in evaluation. A major requirement of this *standard* is to interpret the information obtained in an evaluation against the defensible idea of what *has merit* and *what does not* and the *approaches* revealed. In education, it is possible to make explicit the basis on which the judgements on the worth and merit are based with a much smaller margin of error unlike in development programmes (Boruch & Cordroy, 1980). It is desired that the interpretations of the evaluations should be based on sound basis of perspectives on which the evaluations were addressed and procedures incorporated. Of importance is the rationale for the process of interpretations.

The evaluation documents did not show evidence of descriptions of *perspectives*, *procedures* and *rationale* used to interpret the findings. However, the evaluators were able to defend the interpretations when interviewed. The outcomes are presented in the evaluation Table 4.2.1\_4 below.

**Table: 4.2.1\_4 : Internal Evaluations \_ Descriptions and interpretations**

Two formats were identified. These were:

i) First Format:

- Outcome is spelled out as, *difficulty of topics*.
- Reasons for problems are given, for example, “lack of textbooks”, *inconsistent syllabus*.
- Suggestion for improvement, for example, *provision of textbooks, allocation of more time*.

ii) Second Format:

- The curriculum is broad and overloaded
- There is lack of teaching resources
- There is inadequate in-servicing of teachers

iii) Format for Recommendations:

- Curriculum should be reviewed and revised with the view to reduce and reorganise the content
- Parents responsibility in the provision should be clearly spelt out
- There should be a better system of in-servicing teachers

The outcomes are given against each objective. Where relevant, the findings and the proposed remedy were stated but with no explanation. Refer to the illustration below:

**Table: 4.2.1\_5 : External Evaluations \_ Descriptions and interpretations**

**Sample Objective 1**

To determine the extent to which the objectives of school curriculum are realistic, relevant and achievable.

*Outcome:*

It was found that the school objectives were relevant and desirable but that they were too broad and so not easily achievable.

*Solution:*

There should be a closer look at some of the objectives of some subjects with the view to narrowing their scope and possibly removing some altogether.

*However, the evaluators did not state which ones to remove and why, as well as the implication.*

**Sample Objective 2**

Efficiency and effectiveness of the primary school curriculum implementation process.

*Outcome:*

Awareness creation was not well done, neither was that of teachers adequately carried out. Time spent in the provinces to create awareness was too short.

The *standards* desire that evaluations be made with defensibility of what has merit. To determine if this criterion was met, the researcher analysed the adequacy of purposes laid down for the evaluation, objects of evaluation, units of analysis, methods and techniques of obtaining information, and conclusions as well as the nature of recommendations made. It was established from the reviews that technical adequacy desired in a social

research inquiry was met. The emerging problems were those related to *evaluation methodology* which in most cases were not met.

The person responsible for making the value judgement for the evaluation is generally the one who carries out the evaluation, in this case, the evaluator. However, the interviews revealed that the evaluators were only required to make recommendations based on the findings, but, they were not involved in or responsible for what happened to the results after the evaluation or their utilisation and dissemination. It was only while carrying out informal formative evaluations (that were not large scale) that K.I.E could influence utilisation of results, but, only to some extent and mainly at advisory level. The interviews reveal that issues of implementation and improving teaching skills could be effected but only in guidance towards better skills in the teaching and implementation of the curriculum but not changes of the curriculum content area such as reduction of content and number of subjects taken.

The basis for value judgement of the *Utility Standard* was compromised in a number of ways, taking into consideration a number of *standards*, for example, in the U1\_ *Stakeholder Identification* by not involving all stakeholders in evaluation processes, U2\_ *Evaluator Credibility*, that is, inadequate knowledge in evaluation methodology and U3\_ *Information Scope and Selection*, that is, need for more in-depth inclusion and scope of information. The three *standards* were not met. This means that the basis for value judgment is compromised. It was also established through interviews that the evaluators did not provide alternative basis for interpretation of results, such as debates or alternative

techniques in designing value meanings, such as advocacy. The statements could be referred to as more or less absolute or dictatorial. This resulted in evaluation outcomes being treated with suspicion by the public who were generally not involved in deriving the outcomes.

In order to reach a consensual decision, The Joint Committee recommends that it is useful to solicit the concerns, issues, and clarification because evaluations sometimes arouse controversy. Without consensus the results of the evaluation may not be acceptable to all concerned thus risking its utilisation since decisions on evaluations are often *arbitrary* and are *subject* to debate.

#### **U4 \_ *Standard was Not Met(NM)***

#### **U5 \_ Report Clarity**

According to the Joint Committee, reports should clearly describe the programme being evaluated and may take the form of oral feedback, written memos, video recordings, graphics as well as lengthy documents. For an evaluation to be useful it must be understood. Clarity here refers to explicit and unencumbered narrative illustrations and descriptions.

The titles of the reports depicted the descriptor *evaluation* in most cases. For example, document titles took the following form:



- Formative Evaluation of the Secondary Education Curriculum
- Summative Evaluation of the Primary Education Curriculum.

Various forms of reports were identified:

- Technical evaluation reports
- Summary versions of evaluation extracted from the evaluation reports and bound presenting what appears to be executive summaries.

The reports were written ranging in length from anywhere between thirty-five (35) pages to about one hundred and forty (140) pages. The general outlay of the technical reports is presented in Table 4.2.1\_6a and 4.2.1\_6b.

**Table : 4.2.1\_6a : Evaluation report format**

**Preamble:**

Forward by the director K.I.E

Acknowledgements

Executive summary

Table of Contents

List of Tables

**Chapter One : Introduction**

Background information

Statement of the problem

Objectives of the evaluation

Justification of the evaluation

**Chapter Two: Methodology**

Sample

Instruments

Data collection procedures

Data analysis

**Chapter Three: Findings and Recommendations**

**Bibliography**

**Table: 4.2.1\_6b: Document cover page**

**LOGO**

**K.I.E Research Report Series**

**No**

**Formative Evaluation of the  
Education Curriculum**

**@ Kenya Institute of Education(KIE)**

**P.O Box 30231 \_ Nairobi**

**Date: \_**

An evaluation report as prescribed by the *standards* criteria should carry an executive summary. Not all documents carried an *executive summary*, especially the earlier evaluations that were carried out in late 1980s. Out of the sixteen (16) documents analysed, four did not have *executive summaries*. In later years, from early 1990s executive summaries were included in all the reports. This reflected improvement in evaluation practice over time as pertains to reporting style.

Much as the reports were brief, simple, direct and focused in the manner in which they were presented, they were generally *technical reports* suited for a particularly scholarly audience. There was no evidence of separate summary reports beyond that of a separately bound executive summary. The researcher did not come across any special reports summarised and bound for various levels of stakeholders, policy makers, headteachers, teachers, parents and the community. There was also no translation of the documents for audiences not literate in the language of communication (English). Furthermore, the reports were not made available through open forums to provide explanations or translations to the stakeholders.

As will be seen in later discussions, in A2\_ *Context Analysis*, not enough information was provided to constitute a firm foundation for conclusions and recommendations. The context is adequate for those who have experience in carrying out evaluations and education research in Kenya but not for international evaluators. However, more details in terms of specific aspects of background information were needed. Details related to the curriculum process and principles as they apply to programme content are necessary since

the locals are also not always conversant with all aspects of context within which the programmes being evaluated are judged. This point is further emphasised as exemplified within the U3\_ *Information Scope and Development*.

The results also showed that there was no comprehensiveness in explaining summary statements emanating from the results (see U4\_ *Values Identification*). It was not common to find statements made from previous evaluations or curriculum researches that address the issues found in the report to show continuity and progress into the evaluation process. Even when this was done, the process did not follow up from previously carried out evaluations to show continuity or the reasons why a problem persisted. Furthermore, the documents did not reflect areas for possible future or further study to complement the evaluation as desired by the *standard*. For example, one of the needs assessment studies carried out by K.I.E to support findings of the Koech report (2000) stated the following in the write up:

(P.S: This was in regard to the secondary school curriculum:)

*That“.....In 1990, a curriculum review was carried out leading to revised syllabus. The subjects were reduced to eight from ten and unmanageable areas left out altogether..... The revised syllabi were implemented in 1992..... In spite of the review, problems cited persisted and were reported in the summative evaluation report (1995) which recommended a review of the objectives to make them more realistic and attainable.....”*

The review did not give specifics or details of the referenced findings. By merely stating reduction of subjects from ten to eight without stating which ones or for what reasons they should be reduced, does not give the reader or future reviewers adequate information

on which to act. Other shortcomings were: merely stating that problems cited persisted without stating why; not giving a historical account of an issues derived from previous evaluations; and not identifying the realistic objectives. Without stating what was desirable and realistic leads to lack of commitment in designing evaluation for development.

It can be concluded that much as the documents were well-organised and well-displayed, in-depth information was needed for both pertinent literature and value descriptions of results. Other noted omissions from the reports were failure of descriptive details of procedures and processes on how the evaluation was planned, objectives arrived at, procedures decided on, and the modalities and reasons for the approach chosen. The client or those with vested interest were also not given the chance to review the processes of evaluation. Technical language such as validity, reliability, purposive sampling, were not backed by glossary or separate summary.

#### **U5 \_ *Standard was Partially Met (PM)***

#### **U 6 \_ Report Timelines and Dissemination**

It is desired by the *standards* that the evaluation findings be communicated to intended users at times when information can best be utilised. The intended users are those who commission evaluation, those who are directly affected by evaluation and those who have contributed directly to the evaluation financially.

The extent to which the evaluation reports were circulated could not be established with accuracy. The site visits to the districts, provinces and to the Inspectorate did not reveal existence of the evaluation reports in circulation. This may mean that evaluation reports were either not for circulation among education sectors and officers or if they were, the officers were not willing to reveal their existence. This may also mean that either the reports were not for circulation, a requirement of the central government or if they were, the officers were probably instructed not to avail the evaluation reports. Knowledge of and existence of the evaluation reports were however visible at higher levels of education administration, such as K.I.E itself. This reflects a context where evaluations are prevalently handled in a shroud of secrecy.

The evaluation results at national level were not invariably made public. The results previously made public were those from the needs assessment which were intended for the Koech Report (1999). During the needs assessment, extracts of the results kept appearing in the daily newspapers. The origins of the results were not known but these extracts created a lot of speculation. Releasing parts of the report of on-going national evaluations is common whenever an evaluation is mandated. The problem is that reports appearing in this way are often taken out of context leading to misguided interpretation and misunderstanding of the evaluation results. This could also be a ploy used by disgruntled people such as legislators, to interfere with the outcome of the evaluation results as well as their utilisation.

The guiding principle given in the *standards* is that, a formal agreement should be reached at the outset of the evaluation, that negotiates the authority to fulfill the obligation of disseminating evaluation outcomes and reports. There was no specification of the dissemination process at the outset nor was there a formally agreed dissemination procedure. The terms of the formal agreement to carry out the evaluation reflected the objectives of the study. Elements such as mode of payment were however not reflected.

This discussion shows a tendency of errors made in disseminating results. This relates to releasing reports directly to the client or sponsor, in this case, the M.O.E, the central government and not any other stakeholder. This commonly results in outcomes of evaluations getting treated with suspicion, non-use of results, negative reaction from the public, and lack of control on the use of results.

Another common practice was that evaluations were rushed. Not enough time was given for preliminary preparations and exploratory analysis before embarking on data collection, nor was enough time given for data collection. For example, the length of time taken to conduct an evaluation reflected the following durations.

The field work lasted from October of one year to January of the following year.

**Table: 4.2.1\_7 : Data collection schedule**

Activity	Length of Time
Data collection	3-weeks
Administering teachers questionnaire	2-weeks
Focus groups discussion	November to Mid-January (inc December holidays.)
Data analysis	January

Evaluations do not seem to have been given adequate time. Besides not allocating enough time to carry out evaluations, the results were not used in a timely manner, resulting in a problem of lack of timeliness of utilising results. It is theoretically accepted that for evaluations to have a greater impact, the results should be put to use as soon as the problem is identified. However, it took up to one-year and even longer to *effectively* implement the outcomes of evaluations. To illustrate this point, the researcher gives an account of how advocated recommendations by evaluators have been implemented over time.

The 8-4-4 system of education was implemented in 1985 and first evaluated in 1989.

- The evaluation carried out in 1989/90 recommended the reduction of subjects in secondary schools from ten to eight. The revised syllabus was implemented in 1992
- The problem of overloaded curriculum came up again in the summative evaluation of 1995. It was not until 1999 that suggestions were made in a needs assessment to comprehensively review and restructure the school

curriculum. The changes recommended were finally acknowledged as comprehensive and were finally implemented in 2002 (Daily Nation, 27<sup>th</sup> March, 2002). This was after reluctance by the government.

Two main issues that have come up in all the evaluations were: overloaded curriculum in terms of a large content; too many subjects with not enough time to cover the content; lack of resources for implementing the 8-4-4 curriculum appropriately; and lack of skills by teachers in handling the curriculum. The last two issues are confounded by finances and their implementation could have been hampered due to scarcity of resources typical of government programmes such as those in education. Restructuring the syllabus does not need the same kind of financial involvement but it does require the will to implement and the will to act. It took over ten years to reduce the subjects taught *adequately* and to streamline the course units comprehensively from the time the two issues came up.

This *standard* also deals with other issues, such as, time allowed to conduct the evaluation, editorial work and sensitivity to social impediments of culture, religion and politics. The interviews revealed that internal evaluations were conducted within specified time as there were no deadlines, unlike the comprehensive externally conducted evaluations. However, it was not evident whether formative evaluations were carried out at appropriate times with regard to the programme stage of development and reform as acceptable within evaluation theory and practice. The interviews could not reveal the actual stage at which the evaluations were conducted beyond the fact that the evaluations were indeed carried out.



The *standards* require that the evaluations themselves be timely if impact is to be realised. This was not always the case. Some of the summative evaluations and the needs assessments were generally requested as a demand by the public, and, as a prerequisite for donor aid, and yet still, by legislators as an election issue.

The personnel involved in evaluation at K.I.E also stated that not enough time was given for exploratory measures, such as the needs assessment, deliberation with stakeholders during data collection and at the end of the evaluation study, when the results were ready. They went on to state that this kind of situation resulted in rushed evaluations which may not have been well thought through.

Noticeable also was the absence of controversial issues in the reports despite the fact that the evaluation as practised showed a tendency towards controversies. One of the contentious issues has been the the public's desire to revert the 8-4-4 system back to the previous 7-4-2-3 system (Koech Report, 2000). This issue did not arise in any of the summative or formative evaluation reports.

U6 \_ *Standard was Not Met (NM)*

UNIVERSITY OF NAIROBI  
EAST AFRICANA COLLECTION

U7 \_ **Evaluation Impact**

This *standard* deals with the impact and influence an evaluation has on the decisions and follow-up actions generally referred to as *responsiveness*. It also deals with the influence evaluation has on stakeholders so that they could take into account beneficial actions such

as improving programmes, selecting more cost-benefit products or approaches and stopping wasteful and unproductive efforts.

There was a flaw in this *standard* as there was no indication in the evaluation reports that all interested parties likely to influence responsiveness of the evaluation were involved in planning, conducting evaluation activities and deliberating on the evaluation outcomes. This fact was also established from the interviews. There was no documentation to show that the stakeholders were assisted in assessing and making constructive use of the evaluation results. The only evidence was when interviewing Teacher Advisory Center (TAC) tutors and the head teachers in primary schools. The evaluation information filtering from K.I.E was sent to the TAC tutors who in turn in-service teachers in their areas on issues arising from evaluation. However, the TAC tutors were not certain whether these results were from evaluations carried out for the particular purpose or whether they were from advisory decisions made independently by K.I.E. as they (the TAC tutors) did not seem to have access to the evaluation reports. The reports did not bear any indication either as to whether the stakeholders were informed of the purposes of the evaluation, or the need and use to which the evaluation would be put. Without proper identification of U1 \_ *Stakeholder Identification Standard*, the needs of those with vested interest could not be adequately identified for follow up. This resulted in the evaluation results being rejected or misinterpreted and not utilised, thus hampering follow-up activities that compliment and strengthen programme responsiveness.

The foregoing observations reflect an underestimation of the stakeholders' abilities to react appropriately or in defiance towards the evaluation results. Consider the case of the Koech Report (2000) where defiance of the stakeholders would probably not have arisen if the stakeholders had been involved right from the beginning. This deviates from *standard* recommendation debates and discussions of the evaluation, a process that helps internalise and empower people in owning the programme. Coupled with problems identified in *U6\_Report Timeliness and Dissemination*, there was yet another problem of actualising the impact.

The kind of environment in which evaluation takes place in Kenya as we have seen, is one shrouded with suspicion and secrecy, a system where officers or evaluators could get into trouble should they be open and frank about certain issues, more so, negative ones. If one is to go by the assertion by Taut (2000) of hierarchical governance, this could explain why there were no controversial topics or harsh critical statements dealing with poor performance of the programme or what had not been successful. For example, one does not find a statement that requests for an abandonment of a whole programme but only elements of it.

**Table: 4.2.1\_8 : Decisions taken based on the recommendations**

(Derived from Appendix A, Data source No. 11a)

Decision Taken	Continue %	Modify %	Innovate %	Terminate %
<b>Formative</b>	(3) : 4.8	(6) : 9.8	(3) : 4.8	(1) : 1.6
<b>Summative</b>	(2) : 3.2	(40) : 64.5	(4) : 6.5	(3) : 4.8
<b>Total : (62) : 100%</b>	(5) : 8%	(46) : 74.3%	(7) : 11.3%	(4) : 6.4

**Continue:** Leave things as they are.

**Modify:** Make changes for improvement.

**Innovate:** Introduce something new.

**Terminate:** End activity.

Out of a possible outcome of sixty two (62%) only four (4: 6.4%) components were recommended for termination, forty-six (46 : 74.3%) for modification, seven (7 : 11.3%) for innovation and five (5 : 8%) to be continued.

Evaluations should be reported in ways that encourage use to influence decisions. An analysis was made based on the outcomes and the recommendations to determine the extent to which the manner of reporting encouraged use. This was done by determining whether the outcome statements were *specific* or *generally* stated to influence use and if the outcomes were *speculative*, that is, well thought out or *tentative*. The results of the analysis is illustrated in Tables :4.2.1\_9 and Table :4.2.1\_10

**Table: 4.2.1\_9 : Nature of statement of recommendation made**

(Derived from Appendix A, Data source No. 14)

Information is derived from recommendations arising from formative and summative evaluations.

Statements made	General %	Specific %
Formative	(8) : 12.9	(5) : 8.1
Summative	(15) : 24.2	(34) : 54.8
<b>Total : (62) : 100%</b>	<b>(23) : 37.1%</b>	<b>(39): 62.9%</b>

**General** : Generalises activity to act on

**Specific**: Specifies the exact activity to act on

The results show a situation whereby the outcomes are in most cases specific with up to about sixty two (62.9%) of the outcomes addressing the activity to be acted on. However, too much was left to chance as thirty seven (37.1%) of the recommendations did not address the action to take. Summative evaluation had more generally stated statements than formative evaluations.

**Table: 4.2.1\_10 : Quality of statement of recommendations made.**  
 (Derived from Appendix A, Data source No. 8)

Results derived from outcomes of the evaluations.

	<b>Speculative</b>	<b>Tentative</b>
<b>Formative</b>	(6) : 13.0%	(3) : 6.5%
<b>Summative</b>	(20) : 43.5%	(17) : 37.0%
<b>Total (46) : 100%</b>	<b>(26) : 56.5%</b>	<b>(20) : 43.5%</b>

**SP: Speculative:** Carefully thought about  
**T: Tentative:** Not carefully thought about

The results show that about fifty-six percent (56.5%) of the statements were speculative while about forty-three (43.5%) were tentative. This is almost a fifty-fifty (50:50) chance that statements will be either speculative or tentative.

The two analyses show a situation of reduced impact due to too much generality which means outcomes were not carefully thought about. This may possibly lead to misinterpretation of the outcomes of the evaluation or the clients taking advantage and interpreting the results according to what suits them.

To influence decisions further and to maximise impact, the right users should be accurately identified. This was determined by looking at the targets for the evaluation in terms of those *intended* and those *targeted*. The intended participants were derived from

the statements of the *purpose* of evaluations and the *affected* participants articulated in the *recommendations* made. The results are shown in Table 4.2.1\_11

**Table: 4.2.1\_11 : Users and Targets of Evaluation**  
(Derived from Appendix A, Data source No. 12)

Users and Targets of Evaluation	Intended %		Affected %	
	S	F	S	F
<b>Decision Makers</b>				
Policy_ MOE	27.2	15.0	35.7	35.0
K.I.E	36.4	45.0	29.7	45.0
<b>Recipients</b>				
Learners	9.1	5.0	1.2	1.0
Teachers	18.2	30.0	10.7	15.0
Parents	9.1	5.0	0.0	0.0
Society/Comm.	0.0	-	11.9	-
Schools	-	-	1.2	-
Head teachers	-	-	2.4	-
MOE_Inspectors	-	-	2.4	-
MOE_KNEC	-	-	2.4	-
MOE_TSC	-	-	1.2	-
MOE_TTCs	-	-	1.2	-
Special needs (Handicapped)	-	-	-	-

**S: Summative                      F : Formative**

The results show that evaluation objectives did target policy makers. This was arrived at taking into account the intended and affected targets from both the formative and summative evaluations. The teacher was the second most targeted in both situations. The third most targeted was the school, which is the locale of the evaluand. The learners were more of a target as elements of intended purposes, but, were not the focus while setting the recommendations. This was despite the fact that most of the intended and affected did influence the learner indirectly. In the social science sphere, parents are more of a target

when setting the objectives for the evaluation but not so when stating the recommendations.

The quality of evaluation practice was also determined in terms of the way in which the purposes of study were formatively and summatively stated, that is, to establish the extent to which they were showing immaturity or, showing attainment, respectively.

Table:4.2.1\_12 reflects the results of the analysis.

**Table: 4.2.1\_12 : Orientation of Evaluations**  
(Derived from Appendix A, Data source No. 2)

Nature of Evaluation	Yes	No	Total%
Formatively stated	(11) : 68.8	(5) : 31.2	(16) : 100%
Summatively stated	(4) : 66.7	(2) : 33.3	(6) : 100%

It is important that the criteria for judgement relate to the purpose of the study and the process used in determining the needed information. An analysis was carried out and congruence between objectives techniques and information sought. The results are shown in Table 4.2.1\_13.



**Table: 4.2.1\_13 : Sources of Criteria for Judgment**

Broad Categories	Nature of Criteria for Judgement	Total %
<b>Policy and administration</b>	Curriculum as a programme, school type, classroom size, subject area, learner, teacher, education philosophy relevant of curriculum <b>Other:</b> Enhancing industrial skills	34.6
<b>Process:</b>	Efficiency of implementation, planning, in-service, assess relevance, enhance industrial development	19.2
<b>Outcome measures:</b>	Descriptive : Identified problems, provision reviews, provision of information, provision guidance, make recommendations, feed back information, documentation, teacher competency, develop programmes  Opinion: Attitudes _ school and community Reactions  Student achievement Test scores	46.2
		100%

The results show that:

- emerging elements were relevant to the objects of evaluation and information sought proportionately
- the information sought also related to the key area of focus, that is, the policy, school and to some extent, the community.

This means that there was a balance between the theoretical and practical value, that is, the evaluation having worth and merit, the essence of evaluation. However, due to problems with the U1\_ *Stakeholder Identification*, U3\_ *Information Scope* and

*Development*, and *U4\_ Values Identification*, the criteria for judgement was weakened. The *standards* were only partially met.

As to whether there was failure to respond when results were misused, misinterpreted, and withheld, the results were not really misused or misinterpreted, but withheld. The foregoing discussions allows room for possible misinterpretation of the evaluation results (which is not a good thing) since the recommendations are not always specific and tentative.

This *standard* was therefore not met. The decision was made from the fact that all that has been presented and all that has arisen within the context in which evaluation takes place in Kenya, points to the conclusion that impact seems out of reach as desired by the *standards* criteria.

#### **U7 \_ Standard was Not Met (NM)**

### **4.2.2 Descriptive Analysis of Feasibility Standards**

It is recognised within the *feasibility standards* that evaluations are usually conducted in a natural as opposed to laboratory setting and consumes appreciate resources. As a result, evaluation designs must be practically applied in field settings. Evaluations must not consume more resources than necessary in terms of personnel, materials and time. The

*standards* must therefore be realistic, prudent, diplomatic and economical (Joint Committee, 1994)

### **F1 \_ Practical Procedures**

Evaluators are advised to choose and implement procedures that minimise disruption, as well as feasible and realistic, given the availability of time, budget, personnel and participants. If these *standards* are not adhered to, the evaluation procedures become theoretically sound but unworkable.

The results show that contractual agreements were not drawn while conducting internal evaluations. It was established from interviewing the K.I. E. evaluation staff that the terms of reference were indicated in memo format derived from meetings held in the department or spelt out by the head of the Research and Evaluation Department. Formal agreements were however made between M.O.E through K.I.E and those contracted to carry out external evaluations. The terms of reference generally require that evaluators conduct the evaluations as prescribed, based on the purpose for which the evaluation is called. The length of time in which to carry out the evaluation is given and the cost of the contract spelt out although they are not spelt out in the contractual agreement found in the review documents.

There are a number of steps involved in the development of instruments that should be taken into account in any social inquiry. It was established from the interviews that

piloting was recognised as an important activity. The type of instruments used were those that could easily be developed and adopted for use (Table: 4.2.2\_1).

**Table 4.2.2\_1 : Instruments used**  
 Derived from Appendix A, Data source No. 5a)

Data Collection instruments	Total %
Interviews	44.1
Questionnaire	15.7
Observation schedule	11.0
Check list	9.4
Content analysis	6.4
Document reviews	4.7
Focus group discussion	4.7
Attitudes	1.6
Student achievement measures (Test)	2.4
	100%

Preliminary review of the instruments as well as piloting was deemed important as this helped to determine if they were feasible, realistic and reliable. Piloting was carried out in all cases while undertaken an evaluation. However, piloting may not have always been applied as older instruments can generally be adopted to the new situation with minor changes. There was no need for high level training beyond the expertise of those held responsible for carrying out evaluations. Where necessary the evaluators and the researchers were trained on how to administer instruments and familiarised with the instruments before data collection. However, not enough details were given in the evaluation manifests to enable one to determine how decisions on which instruments to use were made, neither were the recording, storage and analysis of data clarified. This is also recognised in the *Utility\_standard* which was partially met.

As regards data processing, the personnel at the K.I.E are trained and experienced in carrying out the task of data processing and storage.

Concerning the budget and time allocation, no guidelines were given. The K.I.E officers did accept that evaluation were often rushed without consideration of accomplishment of the evaluation task.

It was not easy to determine whether enough financial resources were provided to carry out the activities set out for the evaluation. It was however established from interviewing the K.I.E personnel that at the end of the year, there were some funds left over for evaluations. This could mean that not enough evaluations were carried out. It was not possible to substantiate this statement as there were no records of expenditure in the evaluation documents.

The interviewees revealed that stakeholders were not involved in deciding on instruments for the evaluation, piloting or data collection. Nor were their views on the viability of the schedule for the tasks in hand. This is in contravention of U1\_ *Stakeholder Identification* that requires that the stakeholders are involved at every stage of the evaluation process.

Various instruments are used depending on the evaluator's decision. Interviews and techniques are more commonly followed by checklists and observation schedules addressing physical facilities of schools and actual teaching process in the classroom.

Interestingly student outcome measures in terms of academic achievement and psychological attributes was not popular.

The findings were reported in table format where necessary. They were in the form of descriptive statistics which reflect quantifiable measures such as frequencies, percentages, means and standard deviations. Where analysis of variance was been used, the probability was not shown making it difficult to determine the strength of the significance test. Qualitative information is categorised into commonly occurring interpretation and the percentage of those responding to the reason given. Diagrammatically, the results were presented in table format. Below are samples showing the way in which data was presented:

Modes of presentations of outcome:

**Sample : 1**

Quantifiable information in percentage form in response to the reason why students did not like the secondary school course:

Response	Percentages %
Too much work	30.0
Lack of Textbooks	29.0
Poor teaching	25.0
Lack of teachers	16.0

**Sample : 2**

Quantifiable information, means and standard deviation

Student achievement : mean scores and standard deviation

Subject	Form I	
	Mean	SD
Kiswahili	60	13
Mathematics	18	17.2
Chemistry	16.2	9.2
History/ Geography	49.5	15
Social Education and Ethics	73	16
Home Science	24.6	9
Geography	31.2	13.2

Note:

This was the way the data was presented in the evaluation documents.

**Sample : 3**

In response to the curriculum work load.

Response	Form I %	Form II%	Reason
YES	92.01	84.89	Too many subjects Syllabus loaded
NO	3.05	3.4	--

**Sample : 4**

In terms of in-service carried out?

In-service	Kiswahili %	English %	Biology %	Physics %
None	80	26	44.4	37
1	8.0	15	22.2	29
2	8.0	17	19.4	7
3	4.0	9	16.7	
4	0	5	0	
5	0	11	0	

An analysis of the instruments and the purpose for which they were used, shows that the method is appropriate in meeting the goals of evaluations that are set out, despite inadequacy of U3 \_ *Information Scope and Selection*, U7\_ *Evaluation Impact* standards.

This was established by comparing the purpose for which the evaluation (Table : 4.2.1\_2) was carried out and the match between the purposes of evaluation, the instruments and approaches applied, techniques designed to elicit desired information (Table: 4.2.2\_1). The instruments chosen for the evaluation went a long way in extracting the needed desired to understand the evaluand. This was further strengthened by the scope of information coverage regarding the units of analysis (Table: 4.2.4\_4) which also added to the heterogeneity of the informants.

The *standards* desire that evaluations be carried out at the time when the programme has gone or is undergoing reasonable developmental change. Evaluations should not be carried out whenever it is felt like. Much as the reviewed documents were dated, it was not possible to establish the dates of all the evaluations conducted by the K.I.E. over time because not all the existing documents were availed for the audit trail analysis. Looking at the summative evaluation which was dated 1995, evaluation was conducted ten (10) years after the 8-4-4 system was implemented in 1985. Also the evaluations were based on reviews that were not formative evaluations. This was in 1989. This means that the 8-4-4 did not go through formal evaluation from its inception for a period of three (3) years. This is a flaw since the *standards* require that evaluations be held at periods that are professionally timed. Programmes are generally designed to be monitored constantly, evaluated formatively (halfway), and summatively (when the programme is implemented (completed) as well as monitored.



In Kenya, a national evaluation activity of this nature is centrally managed through the M.O.E. with the head of state as the final authority. As a result public schools are informed of the imminent evaluations and their timing specified without seeking the schools' permission to participate in the evaluation. Fortunately, K.I.E officers and external evaluators are sensitive to the school learning process and so data collection is not carried out during examination periods. Once authority to collect data is given by the government, outside forces cannot stop data collection. On the same note, the same authority can prohibit an individual from conducting the evaluation.

### **F1 \_ *Standard was Partially Met (PM)***

### **F2 \_ *Political Viability***

According to the *standards*, an evaluation has political implications to the extent that it leads to decisions concerning reallocation of resources and influence and its purpose can be achieved with *fair and equitable* acknowledgement of the pressures and actions applied by various interest groups with a stake in it. Evaluation is an inherently political process in that it involves diverse values. There is always the question of who stands to gain or to lose.

The process of evaluation as described earlier and as experienced over the years, reflects the challenges of non-utilisation of evaluation results (see to U\_ *Utility Standard*). The issues arising from the evaluation of the 8-4-4 system of education have been described at length in relation to lack of timely implementation of the recommendations made. On

interviewing the respondents, the outcomes show that the evaluations are not planned with the anticipation of interest groups. This has a direct bearing on the U1 *Stakeholder Identification* standard which was not met. The key stakeholder is the government which has the full control over national evaluations. The government decides what to do with the results and their dissemination and it does not have to explain its actions. The literature on evaluation states that evaluation is power and whoever controls evaluation commands the power. By giving up the responsibility to the evaluator, K.I.E, the donor or the public, the government may feel it is losing control. Other problems arise due to the fact that there is no provision made for the periodic revelation of results by the stakeholders when carrying out externally demanded evaluations for fear of total surprises. In fact, sections of evaluation reports usually find their way into the media whether by approval of the government or not, creating negative debates long before the evaluation report is officially disseminated, thus reducing the chances of the results getting accepted. This also makes the results suspect.

The *standards* criteria require that contractual agreement be made explicitly to govern the evaluation. The format of contracts made between evaluators and the client addressed the purpose of the evaluation. The contracts were referred to as the *terms of reference*. However, the researcher did not locate any contractual letters addressed directly to the evaluators. The information summarised here was derived from the evaluation documents. There were no other conditions given addressing the terms of reference.

As to whether any evaluations have been discontinued and on what account, the answer is yes. Mention was made of suspension of at least one evaluation during interviews with K.I.E. staff. This happened when the 8-4-4 curriculum was being conducted. A directive was issued by the M.O.E. to discontinue the evaluation and it had to be complied with. No reasons were given for this. At the same time the 8-4-4 system of education has been an election issue with political opposition parties demanding its withdrawal altogether.

Much has been raised concerning the stakeholders exclusion in the *U\_Utility Standards*. The right of the public to know has not been upheld and the *standard* is affected by the exclusion of stakeholders at various levels of the evaluation process. This has created a problem in meeting the *standard*.

## **F2 \_ *Standard* was Partially met (PM)**

### **F3 \_ Cost Effectiveness**

According to the *standards* an evaluation should be efficient and produce information of sufficient value so that the resources spent can be justified and that an evaluation is cost-effective if its benefits equal or exceed its cost.

Cost-effective analysis in terms of programme benefits (vs costs) was not considered as one of the objectives in any of the evaluation documents that were reviewed. However, the interviews revealed that money for conducting evaluation was made available whenever an evaluation was commissioned. K.I.E has, in addition, a budget for

evaluation just like other government projected activities. The interviews revealed that the evaluators were able to carry out the evaluation to completion within the projected budget. As to whether the costs were beneficial was not possible to determine as those interviewed could not respond appropriately. Since they are more comprehensive, external evaluations would appear to have been more cost beneficial than the internal evaluation ones.

To determine whether the information produced in evaluations was of sufficient value, issues addressing in *U3\_Information Scope and Selection* were reviewed. Comparison was made between the information source and *U\_7 standard*. The conclusion made was that the information obtained from the evaluations was to a great extent of sufficient value in terms of programme needs but not social needs.

The interviews also revealed that there were no thorough investigations on initial costs of services and materials, neither was there a budget plan in the evaluation documents to lay grounds for credibility to justify this argument. The fact that there were problems with *U1\_Utility Standards* and *Propriety Standards* may imply that any perceived benefits may not be reflected in the costs.

The researcher is of the opinion that the results of the evaluation regarding cost effectiveness is incomplete and needs to be better addressed.

**F3\_Standard was Not Met (NM).**

### 4.2.3 Descriptive Analysis of *Propriety Standards*

*Propriety Standards* are intended to facilitate protection of the rights of individuals affected by an evaluation. They promote sensitivity to and warn against unlawful, unscrupulous, unethical, and inept actions by evaluators. This means that those conducting evaluations should uphold privacy, freedom of information and protection of human subjects. There are eight *standards* in this category.

#### **P1 \_ Service Orientations**

In order to accomplish this *standard*, the Joint Committee states that those who design, administer, use and participate in evaluations must look beyond the self-interest of educators or organizations so as to enhance development of learners and society. Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted recipients, the community and society in general. The application of the *standards* raise a number of issues which, if not addressed or taken into account while planning, designing, conducting evaluation and making recommendations, the evaluation will not serve those it is meant to serve.

In reviewing the reports, the goals of the object for evaluation were not addressed in *totality*. Problems identified in U\_ *Utility Standards*, particularly U1\_ *Stakeholder Identification*; U4\_ *Values Identification*, U6\_ *Report Timeliness and Dissemination*, U7\_ *Evaluation Impact*, do affect service delivery. Much as the key persons directly affected and influenced by evaluation of curriculum were identified, that is, teachers,

learners and the immediate community, the *standards* require that participants be well served, as well as the community and the society. The targeted users as derived from the recommendations were more often policy level decision makers who made decisions on the programme in general than learners and teachers. The society was not targeted. The results are reflected in Table 4.2.4\_4.

The *standards* also require that the recommendations must target the users and those influenced by evaluation results and this must be reflected in the statements of the recommendations with clarity and directness. An analysis was carried out to determine the clarity and directness of the recommendations made and those targeted in terms of specifics and generality. The results in Table 4.2.1\_9 show that thirty-seven percent (37%) of the time, the recommendation statements are generalised. The following shows examples of specific and general statements that were made:

*Specific statements:*

- Teachers should be trained in *testing in education*
- There is need to expand K.I.E to develop school text books
- Teachers work load should be reduced from 45-50 to 30 lessons
- There should be no extra tuition after class hours

*General statements:*

- Curriculum overload should be reduced.... *what load? .... and reduced in which way?*
- There is need to remove unnecessary overlaps .... *which overlaps?*
- Content area should be improved .... *which content area? .... and improved in what way?*
- Enforce policy on the number of subjects required.... *does this mean reduce or increase number of subjects?*

Questionable was whether the promised services were delivered. This arose as a result of problems with the U6 *Report Timeliness and Dissemination* standard which was not met due to the length of time it took to implement evaluation results. The reviews revealed that it took up to ten-years to implement major programme concerns identified during evaluations.

It was also difficult to determine the extent to which programme effectiveness was monitored. The interviews revealed that there were no designed systematic approaches that were familiar to educationists and the school personnel as they relate to monitoring the curriculum desired by the *standards*. The Inspectorate is the other organ responsible for *policing and monitoring* the curriculum directly (M.O.E, 1987). The interviews further revealed that the inspectors visit schools as a duty to oversee programme implementation by the school administration and teachers. The curriculum was seldom the core business of the inspectors, nor were problems arising from the curriculum or the difficulties arising from implementing the curriculum part of the inspector's mission during the visits. The teachers claimed that their views on curriculum issues were not sought and that the visits were rushed giving an impression that the exercise as carried out as a routine check on teachers and not as an essential function of the evaluation exercise. This shows that much as monitoring is recognised as an activity for programme evaluation, it is not planned or executed according to evaluation principles.

Metaevaluation, the essence of this study, is a desirable function of the evaluation process by the Joint Committee. It was however not part of the evaluation process at any level. This denies the evaluators an opportunity for self-assessment.

The reviews reveal that there have been no recommendations to terminate a programme in totality after an evaluation, only in segments. There were recommendations that requested for termination of segments of the programme, such as, pastoral course, extra tuition and school levies.

There have been cases whereby programmes had been discontinued due to government directives. The interviews reveal that *New Mathematics* was discontinued and then later reintroduced. Yet, there was no evaluation done to justify the need for discontinuation or reverting back. The initial outcry opposing *New Mathematics* came from the public, the politicians and the professionals.

Another concern arising from the audit trail was that matters pertaining to the evaluation process and procedures have not lived to the state of the art of evaluation. It is evident that the approach taken by the evaluators, both internal and external, in conducting the evaluations, emphasises research methods and designs that pertain to *A\_Accuracy Standard*, which is, the technical aspect of social research approaches, much more than the evaluation processes. There is, therefore, inadequate attainment of various *standards* in terms of evaluation methodologies. Reference is made to issues arising from the



*U-Utility Standards.* These have an influence on the extent to which service delivery can be achieved.

### **P1 \_ Standards was Partially Met (PM)**

### **P2 \_ Formal Agreements**

This *standard* obligates the formal parties to an evaluation so as to agree in writing as to what is to be done, how, by whom, when it should be done so that the parties adhere to all conditions of the agreement and renegotiate them.

The only time a formal contract was entered into was when an evaluation was commissioned by M.O.E through K.I.E with external evaluators. It was established that no agreement was made when carrying out internal evaluations, only a brief memorandum of understanding on the terms of references was made while conducting evaluation. Education officers in various provinces and districts were notified of the evaluation, but, whenever an agreement was made, it was not comprehensive, although there was no statement of actions to be taken should a breach of contract occur, or how to handle a dispute. No single breach of contract was recorded in the reports or established through interviews. Only two agreements were found in the summative reports where they were also referred to as terms of reference. The terms were based on the purposes and objectives of the desired evaluation.

What was to be done, how, by whom and when was not stated in the evaluation reports. There was also no mention of a penalty should there be a breach of contract or what a breach of contract actually was.

## **P2 \_ Standard was Not Met (NM)**

### **P3 \_ Rights of Human Subjects**

The *rights* deal with the aspect of evaluations being designed and conducted to respect and protect the rights and welfare of human subjects.

The interviews reveal that the evaluators were aware of the regulations that pertain to conducting evaluations from their experience in social science research. Interviews revealed that considerations were made regarding the respondents. There were short-comings, though. For example, consent to carry out evaluations in public schools was not sought nor was permission sought from student participants in the evaluation. Directive for a school to be part of an evaluation was issued by the M.O.E and the school was not given an option to refuse, but had to go along with the exercise with no option of withdrawing. The parents were not consulted if their children were sampled for the evaluation. There was no information on what was intended. However, the researcher noted that the nature of data collected and the kind of instruments used did not require long involvement of participants to raise concern, neither did the methodologies applied put the participants in *harmful* and *uncomfortable* experiences. The instruments generally

included basic achievement tests, interview schedules and questionnaires for teachers and students (Table: 4.2.4\_3).

However, some considerations were made to meet various needs related to the *standard*.

For example, privacy of information was achieved without requesting for the name of the respondent on the instrument. The evaluation and teaching periods were respected.

The standards further required that evaluators understand various values related to cultural and social values of participants. The evaluators were aware of this although such cases did not arise due to the mode of inquiry and the instruments used which did not compromise the personal beliefs of the participants.

### **P3 \_Standard was Not Met (NM)**

#### **P4 \_ Human Interactions**

There is need to respect human dignity and worth in interactions with other persons associated with an evaluation so that participants are not threatened or harmed.

This *standard* is influenced by the UI\_ *Stateholder Identification* due to the nature of the approaches and techniques used in acquiring information. The participants were not exposed to harmful situations nor were their human dignity violated at any time. None of the groups were more privileged than others nor were findings on participants biased. There was no revelation of results of participants. However, views and feelings

participants were not considered while carrying out evaluations, nor were there open communication channels with the evaluators.

#### **P4 \_ *Standard was Not Met (NM)***

#### **P5 \_ Complete and Fair Assessment**

The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the programme being evaluated so that the strengths can be built upon and problem areas addressed. This *standard* does not refer to generating an equal number of strengths and weaknesses. It means that the evaluator should be thorough and fair in assessing and reporting both negative and positive aspects of the programme.

The *standard* relates to A11 \_ *Impartial Reporting* which addresses all perspectives of and U5 \_ *Report Clarity* which deals with presentations of reports at all administrative levels, with firm foundations of methodology, conclusions and recommendations.

The information was not significantly comprehensive to allow fair assessment of the reports as presented. Not all perspectives of stakeholders were addressed, nor were the stakeholders' views sought when reviewing the results and recommendations. The views were based on the evaluators' perspective. The methodology was reported in both the internal and external evaluations. The limitations in conducting evaluation were not comprehensively addressed. They addressed limitations equated with social science inquiry.

To ensure fair assessment of the evaluations, the findings were analysed to determine if the reporting was balanced in terms of positive and negative representations of the results.

The analysis is presented below:

**Table: 4.2.3\_1 : Positive and Negative presentation of Outcome**  
(Derived from Data source No. 8)

	-ve %	+ve %	Neutral %
<b>Formative</b>	(8) : 22.9	(0) : 0.0	(1) : 2.8
<b>Summative</b>	(11): 31.4	(3) : 8.6	(12) : 34.3
<b>T : (35) : 100%</b>			
	<b>(19) : 54.3</b>	<b>(3) :8.6</b>	<b>(13) : 37.1</b>

Indeed, the positive and negative aspects of the evaluation are presented, however the reporting tended to address more of the negative aspects of the results.

#### **P5 \_ Standard was Not Met (NM)**

#### **P6 \_ Disclosure of Findings**

This *standard* deals with the fact that formal parties to an evaluation should ensure that the full set of the evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with express legal rights to receive the results.

There was a problem in attaining this *standard*. The results revealed that the technical report was produced for the M.O.E through K.I.E. and it was not circulated to interested and affected parties such as education officers, education experts and school personnel or the community. The summaries of the evaluation reports generally reflect the executive summary of the evaluation reports themselves. This only came up in the late 1990s. There was no oral reporting of results nor interim reporting. In fact the whole issue of findings is generally shrouded in secrecy. Under these circumstances the persons and groups concerned do not have uncontrolled access to the results except the K.I.E personnel. The K.I.E personnel were also not at liberty to issue evaluation reports to those needing them for purposes of research or general information.

The tendency of non-disclosure of findings has threatened evaluator credibility as the evaluators are not involved in decisions of how to handle the evaluation outcomes and impact. If the persons and groups who are affected by the evaluation results cannot get the results and information about the evaluation, then according to the *standard* guidelines, they cannot make constructive *use of* the evaluation results.

The problems associated with this *standard* are as a result of inadequate attainment of other standards such as, P2 \_ *Formal Agreements*, which was lacking in details whereby there was no assurance of compliance with the right to know, when and how the results are to be released and used; A2 \_ *Context Analysis*, which requires that the programme be examined in detail so that its likely influence is identified; U6 \_ *Report Timeliness and Dissemination*, which deals with different kinds of report formats beyond the technical

report as well as the dissemination plan. There was no agreement on a dissemination plan prior to the evaluations.

It is also desirable that the findings should be *useful* so that the strengths of the programme will be built upon and problem areas addressed. The key term here is usefulness. The element of usefulness of the findings was analysed by looking at the format in which the results were presented, nature of the recommendations made and how they were reported, significance of scope of information and how significant the information was to the stakeholder. Most of these issues also fall within the

*Utility\_ Standards and Accuracy Standards*

The analysis involved determining general and specific statements in terms of the extent to which they were action oriented. The results given earlier show that there was more generality of activity and action on the specific and exact actions addressed in the recommendations (Table :4.2.1\_9). There was a tendency to present the recommendations in specific terms, full (62.9%), than in general terms (37.1%). The usefulness of the results were further determined by looking at the nature of the statements of recommendations made in terms of their speculativeness or their tentativeness. The results show that outcomes tend to be more speculative (56.5%) rather than tentative (43.5% ; Table: 4.2.1\_10). As to whether purposes of various categories of evaluations were stated to define the *orientation* of evaluation, sixty-eight (68.8%) were formatively stated and sixty-six (66.7%) summatively stated (Table 4.2.1\_12). This means that not all

purpose statements define the evaluation function adequately to give direction of the evaluation activity so as to enhance nature and disclosure of findings.

Much as the results are useful and could be built upon and problem areas addressed, the *standards* criteria were not met.

#### **P6 \_ *Standard* was Not Met (NM)**

#### **P7 \_ *Conflict of Interest***

Conflict of interest should be dealt with openly and honestly so that it does not compromise the evaluation process and results. The concern is that the *personal* or *financial* interests of an evaluator might either influence an evaluation or be affected by it.

A major conflict in the evaluation concerns enterprise has been on programmes that are pro-education. In the case of Kenya, personal interest in claiming ownership of the programme by the government has been evident while the opposing parties have taken a view which is against the 8-4-4 system of education. As to whether the opposing non-governing party is justified is difficult to determine. The recently evaluated 8-4-4 system of education brought to surface conflicts by all interest groups, including the community. The 8-4-4 system of education was also an election issue in the 1997 general election. The popular view was to get rid of the current 8-4-4 system and to revert to the previous 7-4-2-3 system. This was resisted by the government.



The internal evaluations and the previously commissioned evaluations were not published and so they did not create much public interest and in most cases the public was not even aware of the ongoing evaluation. Conflicts were however not dealt with openly and honestly, nor were possible sources of conflicts mutually acknowledged. Concerns were also raised about involvement of the stakeholders, because stakeholders they had not been appropriately involved. With the *U1\_ Stakeholder Identification* and *U6\_ Disclosure of Findings* not met, this would have a direct bearing on achieving this *standard*.

This discussion shows that evaluations have a potential for conflict and the issues should probably be based on the idea of not avoiding the conflict but how to deal with it.

There was no conflict regarding the external evaluators commissioned to carry out evaluations regarding evaluator bias. This was because there were only a few external evaluations. In each case, there was a deliberate move to use different evaluators each time. The issue of co-opting did not therefore arise as there was no problem of an evaluator pleasing the client to gain favour for a possible chance of carrying out other evaluations in the future. This helped minimise the possibility of falsifying and creating possible bias in the processes used, the findings and the interpretations. It was not evident as to whether procedures to protect against conflicts were given in the agreements.

Other related *standards* that had a bearing on this particular *standard* are P2 \_ *Formal Agreements*, whereby agreements are not agreed upon in writing; F2 \_ *Political Viability*, resulting in being aware of losses in terms of monetary gains as well as social, moral and political leverage by various groups and to prepare to resist ensuing pressure. These *standards* were not met.

#### **P7 \_ Standard was Partially Met (PM)**

#### **P8 Fiscal Responsibility**

This *standard* deals with expenditure in terms of operational costs of evaluations.

There were no records on funds of any kind in the evaluation reports. It was therefore not possible to determine accountability in terms of allocation or expenditure, nor could one determine if funds provided were used for the purpose and procedures stated in the evaluation. Those interviewed were not ready to discuss fiscal matters of K.I.E or M.O.E.

#### **P8 \_ Standard was Not Met (NM)**

### **4.2.4 Descriptive Analysis of Accuracy Standards**

The *standards* in this category determine whether an evaluation has produced sound information. The Joint Committee states that the evaluation must be comprehensive. Many of the programmes' identifiable features should be considered and data on those

particular features should be gathered on those particular features judged important. The information must be technically adequate, and judgements rendered must be linked logically to the data.

### **A1 \_ Programme Documentation**

The programme being evaluated should be described and documented clearly and accurately to gain sound understanding of the programme in terms of the nature and implementation.

Two issues arise here. The reports reflected were documented clearly but not comprehensively according to the *standard* requirement. However, external evaluations presented more details on the historical development of education reports. U3\_Information Scope and Selection has a bearing on this *standard* yet it was only partially met. It was as desired in the U3\_Standard that more thought needed to be given to deriving information scope and programme development of evaluation knowledge from one evaluation to the next. Both internal and external evaluations presented the objectives articulately although their coverage was not adequate. Other information desired by the *standards* was not relevant as the evaluations did not reflect measures that posed risk to the participant, such as potential extreme side effects.

Characteristics such as personnel involved in the evaluand (or the curriculum), costs, and procedures used in managing the programme activities are important. These were

however not addressed. The *standard* requires that reference be made to various characteristics of the programme being evaluated. Characteristics such as personnel, costs, procedures in implementing the programme, location, facilities, the setting and comprehensive nature of participants were not given, nor were there separate descriptions for each aspect of the programme under review.

This reflected a shortcoming on meeting the need to gain sound understanding of the programme in terms of the nature and implementation.

#### **A1 \_ *Standard was Partially Met (NM)***

#### **A2 \_ *Context Analysis***

The *standards* maintain that the context should be examined in detail so that the likely influences on the programme can be identified. This includes geographical location, social climate, competing activities and any economic conditions.

The evaluation documents did not address any of the issues listed above to provide sufficient information to aid in understanding the implication of each attribute. It is acknowledged that the political climate can be tense when evaluations are influenced by donors, the public or political opposition parties. The F2\_ *Political Viability* standard was not met.

There was need to discuss the hierarchical system of governance which generally reflects the managerial type of evaluation whereby the evaluators have no control over what they do, what the evaluation objectives should be, and what should happen to the evaluation results. There was also need to present the issues that have plagued the 8-4-4 system of education, such as costs and curriculum overload and the fact that there has been numerous calls to get rid of the 8-4-4 system altogether (Daily Nations, 27<sup>th</sup> March, 2000). The interviews revealed knowledge of this situation but those writing the evaluation reports did not address the said issues claiming that they were sensitive and would not augur well with the government of the day.

Information on the economic situation would be of importance as it helps in understanding reasons why certain actions are either taken or not taken. Cost elements do have direct influence on resources which are almost always inadequate, teachers in-service which was adequately carried out, and the limitation on procedures that may affect practicality. The *A\_Accuracy Standard* deals with the technicality aspects of research methods and design. The cost element also has an effect on the extent to which changes recommended could be implemented, such as those involving reverting from 8-4-4 system to 7-4-2-3 system of education. The sources from which the evaluators got their information are listed below:

- Kenya Government Development Plans
- Working Party Reports on Education
- Education Commission of Inquiry Reports
- Sessional Papers on Education
- UNICEF \_ materials, for example, *State of the Worlds Children*
- Ministry of Education Curriculum Syllabus \_ Primary and Secondary
- Journal Articles on Education research

No reference was made to the current trends of debates in the field of interest, that is, education in general and the curriculum in particular. For example, Education For All (EFA) needs, gender issues in education, careers and the curriculum were neither presented in the evaluation documents nor addressed in the evaluation.

## **A2 \_ *Standard was Partially Met (PM)***

### **A3 \_ Described Purposes and Procedures**

The *standard* requires that the purposes and procedures should be monitored and described in sufficient detail so that the points of agreement and disagreement are identified and assessed.

The extent of descriptiveness of purposes and procedures were found to be inadequate. The purposes of evaluations were stated in terms of evaluation objectives but not intended use of the results as desired by the *standards*. This was the case in all the evaluations whether formative, summative, or needs assessment. Table 4.2.4\_1 shows a sample of commonly addressed objectives:

**Table: 4.2.4\_1 : Sample of commonly addressed objectives**

Derived from Appendix F, Data source No. 2

<b>Purpose of evaluation</b>
<b>Development and Improvement, for example:</b> To recommend procedures for strengthening education in Kenya To make recommendations for school improvement
<b>Awareness and knowledge, for example:</b>
<b>Formative</b> To determine the extent to which the school curriculum achieves objectives of school To determine the extent to which the school curriculum has been effectively implemented
<b>Summative:</b> To determine the extent to which education objectives are realistic and achievable To assess the relevance of the school curriculum
<b>Needs Assessment:</b> To provide information on national goals of education To provide information on curriculum content for schools

The interviews revealed that the purposes of the evaluations were not described, evaluated or reviewed at any time during the evaluations. Also, the intent of the evaluations was not given much as it could be derived from the objective statement as presented. The *standards* require that details should include description of purposes and procedures in the executive summary and a full technical report. This was not evident and is illustrated in *US\_Report Clarity*. It was established that there was failure on the part of the evaluators to provide descriptive details on procedures and processes on how evaluation was planned, objectives arrived at, and procedures decided upon. The procedures used to carry out the evaluations were not described in detail and could only be adequate to those who have knowledge or expertise in social research methods.

**A3 \_Standard was Not Met (NM)**

#### A4 \_ Defensible Information Source

It is required by the *standards* that sources of information used be described in sufficient detail, so that the adequacy of information can be assessed.

Much as there was an attempt to include a variety of sources in determining the information for the evaluations, the sources were not described nor were the reporting, criteria and methods used to decide on their choice addressed in the reports as part of the documentation. Sources from which information was derived are given below and they are general and appropriate for this kind of evaluand.

**Table: 4.2.4\_2 : Sources of Information**

(Derived from the interviews and evaluation documents)

<b>School Level</b> Students Teachers Parents Head Teachers BOG School Records Physical Facilities Classroom learning  <b>Source : 16 review documents</b>	<b>MOE</b> Education officers Inspectorate K.I.E K.N.E.C T.S.C K.N.U.T	<b>Out of School</b> School leavers: Dropouts:	<b>Other</b> Education scholars Community leaders Religious leaders

The most popular mode of eliciting information was interviews (44.1%) and questionnaires (15.7%) for teachers, pupils and parents. However, certain information from important perspectives was missing. These were classroom interaction, curriculum course content analysis, gender analysis including analysis of programme costs.



**Table: 4.2.4\_3 : Instruments used to elicit Information**  
 (Derived from Data source No. 5a)

<b>Data Collection instruments</b>	<b>F</b>	<b>S</b>	<b>NA</b>	<b>Total %</b>
Interviews	18	26	12	44.1
Questionnaires	12	4	4	15.7
Observation Schedules	8	4	2	11.0
Check lists	6	2	4	9.4
Content Analysis	2	2	4	6.4
Document Reviews	-	2	4	4.7
Attitudes	-	-	2	1.6
Student outcome measures	2	-	1	4.7
<b>Source =The sixteen evaluation review documents</b>				

Sampling units of analysis do create a variety of complexity of sources by categorising units of analysis and defining crosscutting issues. These were determined from the evaluation reports and represented in Table 4.2.4\_4.

**Table: 4.2.4\_4 : Units of Analysis**

(Derived from Appendix F, Data source No. 6)

Unit Categories	Cross Cutting Issues	Cross cutting issues
School Types:	<b>Formative</b> Day, boarding, mixed, single, national	<b>Summative</b> Day, boarding, mixed, national, boys / girls, Year of establishment
Region:	Rural/urban	Urban/Rural; province/ district;
Class levels:	Grades, sizes, learning process	Grades, sizes, learning process, gender
Course units:	Sequence, context, time allocation, relevance to objectives, process, Text books	Sequence, context, time allocation,
Attitudes:	Towards school, 8-4-4	Towards school, 8-4-4
Respondents: -Teachers	Teachers, head teachers' Learners, Community	Province, gender, age, academic and profession
-Head teacher	--	Province, gender, sex, age, academic & profession

The analysis shows that the units present a variety of components that categorise the context within which schooling takes place, thus making them appropriate.

The internal evaluations are characterised by sources, such as school level, individuals, mostly teachers and students. Parents are included depending on the design of the evaluation. Not included were the education officers from other M.O.E departments. Sampling was done in both cases and included crosscutting issues such as region (rural/ urban) and gender.

Much as it was possible through the analysis to extract the information and to ascertain justification of what was done, there was not enough *information* provided to adequately assess the situation.

#### **A4 \_ Standard was Partially Met (PM)**

#### **A5 \_ Valid Information**

This *standard* deals with the extent to which the interpretations deduced are valid for the intended use. *Validity* as used here concerns soundness and trustworthiness of inferences made from the results. The congruence is analysed by relating the information sought and the objective of the evaluation.

To determine the congruence of the information sought and the stated objectives, the documents were reviewed looking at the objectives as described and the information sought. This was done by comparing the purposes of evaluation in Table : 4.2.1\_1 and the extent to which the purposes of the evaluations were addressed in Table: 4.2.4\_5 The researcher thought it worthwhile to compare the two attributes due to the claim made in U3\_ *Information Scope and Selection* that there was a tendency of commonality on the issues addressed that despite the differing needs of evaluation, it almost always related to the following: objectivity and relevance of the school curriculum, issues of implementation of 8-4-4 curriculum, resources and facilities available, teacher quality and competency (see to *Utility standard*). Table 4.2.4\_5 represents the extent to which the

statements of the outcome of evaluations as derived from the documents reflect the object of evaluation proportionately.

**Table: 4.2.4\_5 : Extent to which outcomes address object of evaluation**  
(Derived from Appendix A, Data source No. 9a)

Object of Evaluation	Coverage Of purpose %	Extent to which outcome reflects purposes%
Education philosophy	38.7	10.7
Curriculum programme	29.0	28.6
Subject area	12.8	7.2
Personnel _teacher	6.5	17.2
Facilitator _also Teacher	6.5	21.4
Reactions	6.5	7.1
Policy level		7.8
Learners needs _ indirectly		
	100%	100%

The results show that there was no congruence in a number of pairs. That is, coverage of purpose and extent to which outcome reflects purposes. Consider *education philosophy*, 38.7% : 10.7%; *subject area* 12.8%: 7.2%. Congruence was noted between the curriculum programme which gave a comparison figure of twenty-nine (29%) coverage and Twenty-eight (28.6%). The idea here was to show whether the results as interpreted reflect in equal proportions the object of inquiry or a balance. The results also show that there was no analysis in the evaluations to determine congruence of the results in the manner presented in this research.

The issues of inadequate comprehensive description of various segments, procedures and judgmental process has been cited in the research in that there was no appropriate description to enable sound and trustworthy validation of the on-going activities. The desired constructs and behaviours were also not described. There was also no validity claims sought neither was the context within which the evaluand was found adequately addressed. See *A2\_ Context Analysis standard* which was not met.

The *standards* require that evidence of information be presented in both qualitative and quantitative measures that justifies their use. There was evidence of more quantitative than qualitative measures sought as presented in the techniques for information seeking in Table: 4.2.4\_3.

In most instances, it was not possible to address the objectives of the evaluation without jeopardising the needed outcome. Of course, the results could have been more detailed and the validity increased if other perspectives were included (See *A4\_ Defensible Information Source*). The validity of information depends on *A2\_ Context Analysis* which guides the nature of information sought. The *A2\_ standard* was only partially met.

A number of flaws were noted. These included, lack of and inadequate description of constructs used, procedures of implementation which arose as a result of not meeting *U4\_ Values Interpretation* standard. There was no overall assessment due to the fact that *F3\_ Cost Effectiveness standard* was not met, and so it was not possible to determine fully the degree of evaluation questions and answers.

As stated earlier, some validity is ensured by the fact that the personnel for evaluation were assessed as credible. Consider U2\_ *Evaluator Credibility* standard which was met. The characteristics of the respondents were also considered appropriate.

#### **A5 \_ *Standard Partially Met (PM)***

#### **A6 \_ *Reliable Information***

The issue arising here is the extent to which the information obtained from the data gathering system is consistent. Consistency of information is affected by random errors in each procedure.

There was no stated reliability of the tests to determine academic performance or the other scales of measure used. As regards administering instruments concerning open-ended interviews, no systematic procedure was evident to determine rater reliability. The *standards* criteria encourage discussions on developing prepositions, interpretations and conclusions by two or more impartial peers to be held in clarification of one's point of view. The decision of the interviewer or evaluators in all cases was final and there was no evidence of discussions involving the stakeholders and the evaluators. In summary, the rationale in relation to the procedures, administering instruments, and the heterogeneity of the group was not provided. This made it difficult to determine distinctly the appropriateness of this *standard*.

However, the scores and analysis were assessed as credible in relation to *technical* attributes but not *evaluation* attributes (see U2\_Evaluator Credibility standard).

#### **A6 \_ Standard was Not Met(NM)**

#### **A7 \_ Systematic Information**

This *standard* implies that all information that exists or is new is free from error and is kept secure.

There was no information given in the reports to show the methods used to control error arising from data collection, data scoring, recording, coding, analysing and reporting. With the A6\_ *Reliable Information standard* not met, and the P1\_ *Service Delivery standard* not met, it was difficult to determine the adequacy of this A7\_ *Systematic Information*. There were also no systematic accuracy checks or programme training processes. Much as data analysis was carried out by data specialists, there was no evidence of an outsider involved in verifying the data, nor was there a plan to verify the data process. This creates a problem.

The errors commonly found by the evaluators in the evaluation documents were those arising from data collection. Evaluators hoped that the respondents were honest in their responses and that they understood the questions.

#### **A7 \_ Standard was Partially Met (PM)**

## **A8 Analysis of Quantitative Information**

This *standard* should be systematically analysed so that evaluation questions are effectively answered.

The *standards* require that quantitative information addresses socio-economic characteristics, measures of achievement, attitudes, behaviour descriptions, and, materials being evaluated. An illustration has been given on the extent to which this is addressed.

Social-economic characteristics and behaviour were not evident. The quantitative measures used were appropriate, consider *A4\_Defensible Information Sources and A5\_Valid Information*, but limited. The analysis was found to be systematic but lacked a process involving exploration as well as complex analysis. The visual displays were presented using tables mainly. The reports were devoid of graphs and pie-charts.

### **A8 \_Standard was Partially Met (PM)**

## **A9 Analysis of Qualitative Information**

The information should be appropriately and systematically analysed so that evaluation questions are effectively answered.

Qualitative information arose from interviews with parents, board of governors, and community groups (Table : 4.2.4\_2). There was also content analysis. The standards



require appropriate analytical procedures such as summarising data. The analysis was presented as categories of information in form of a percentage reflecting group representation. There was a short-coming identified in *A6\_Reliable Information* as it was established that an inter-rater reliability was not one of the procedures. This could reduce the reliability of the information.

### **A9 \_ Standard was Partially Met (PM)**

#### **A10 \_ Justified Conclusions**

It is recommended within the *standards* that conclusions reached in an evaluation must be defensible and assessed by the stakeholders. Without adequate information to determine this, an evaluation may be discounted.

Generally, there are no justifications of the results, neither are alternative explanations of the findings given. However, the conclusions are limited to the situations, time period, persons, context and purposes for which the evaluation is applicable. This *standard* requires that feedback be provided by programme participants about credibility of interpretations, explanations, conclusions and recommendations. Throughout the evaluation process, the participants were only included in their capacity as respondents during data collection. This resulted in conclusions reflecting only the evaluator's views. The needs assessment required for the Koech report(1999) was the only instance where conclusions and recommendations were discussed with a panel of experts. Even then the discussions did not create much influence on the government to change its stand on

certain pertinent issues. The government, however, did finally reduce the number of subjects showing that discussion of results in this manner goes a long way in influencing use of results.

#### **A10 \_ *Standard was Partially Met (PM)***

#### **A11 \_ *Impartial Reporting***

Generally, reports on evaluations tend to be distorted in a number of ways. The *standards* therefore, recommend that reporting procedures guard against distortion caused by personal feelings and biases.

Firstly, the reports did not reflect all the perspectives that should be taken into account, for example, U2 \_ *Values Identification* standard that was not met. There was also a problem arising from the fact that it was the owners of the programme who carried out ninety (90%) of the evaluations especially internal evaluation. Agreements in carrying out the evaluations were reached, but no agreement was reached as to the steps to take to ensure the fairness of the evaluations. There was no evidence to show any attempt to establish and maintain independence. None of the approaches reflected adversary-advocacy, outside audit or even externally conducted evaluations. There was no description of the steps taken to protect the integrity of the reports, and there is no involvement of the public in open presentations during planning, data collection and final reporting.

**A11 \_ *Standard was Not Met (NM)***

**A12 \_ *Metaevaluation***

Metaevaluation was not carried out.

**The Standard was Not Met (NM)**

### 4.3 Performance of metaevaluation on *Standards\_Criteria*

**Table: 4.3.1a : Summary of Metaevaluation results**

DESCRIPTOR	Met	Partially Met	Not Met	N/A
<i>Utility Standards</i>				
U1_ Stakeholder Identification		X		
U2_ Evaluator Credibility		X		
U3_ Information scope and selection		X		
U4_ Values Identification			X	
U5_ Report Clarity		X		
U6 Report Timeliness and Dissemination			X	
U7_ Evaluation impact			X	
<i>Feasibility Standards</i>				
F1_ Practical Procedures		X		
F2_ Political Viability		X	X	
F3_ Cost Effectiveness			X	
<i>Propriety Standards</i>				
P1_ Service Orientation		X		
P2_ Formal Agreements			X	
P3_ Rights of Human Subjects			X	
P4_ Human Interaction			X	
P5_ Complete and Fair Assessment			X	
P6_ Disclosure of Findings			X	

P7_Conflict of Interests		X		
P8_Fiscal Responsibility			X	
<i>Accuracy Standards</i>				
A1_Programme Documentation			X	
A2_Context Analysis		X		
A3_Described Purposes and Procedures			X	
A4_Defensible Information Source		X		
A5_Valid Information		X		
A6_Reliable Information			X	
A7_Systematic Information		X		
A8_Analysis of Quantitative Information		X		
A9_Analysis of Qualitative Information		X		
A10_Justified Conclusion		X		
A11_Impartial Reporting			X	
A12_Metaevaluation			X	

**Table : 4.3.1b : Summary of metaevaluation on *Standards\_Criteria***

(M) :Met ; (PM) : Partially Met ; (NM) : Not Met ; (NA) : Not Met

<b>Standard Attributes</b>	<b>(M)</b>	<b>(PM)</b>	<b>(NM)</b>	<b>(NA)</b>	<b>Total</b>
<i>Utility</i>	-	4	3	-	7
<i>Feasibility</i>	-	2	1	-	3
<i>Propriety</i>	-	2	6	-	8
<i>Accuracy</i>	-	7	5	-	12
<b>TOTAL : 30</b>	<b>0 (M)</b>	<b>15(PM)</b>	<b>15(NM)</b>	<b>0 (NA)</b>	<b>30</b>

## **CHAPTER FIVE**

### **SUMMARY, CONCLUSION and RECOMMENDATIONS**

In this chapter the researcher sought to create a convergence and to address the purpose of the evaluation which was to determine programme evaluation practice in Kenya and to hopefully guide improvement of evaluation practice and process through the recommendations.

#### **5.1 Summary of the Evaluation Research Results**

##### **5.1.1 The evaluation context**

Issues of power and programme needs are central to understanding how evaluation systems work in Kenya and not evaluation methodological principles. It is not possible for an accounting person or body to give an account of its activities while conducting evaluation without expectations that it will be subject to some form of constraint or sanction should the account prove to be unsatisfactory to the government or the bureaucracy. The accountable relationship is between the following bodies:

- the government of the day
- the opposition political parties
- the donor community
- the general public
- scholars in the field of interest

The frames of the ensuing debates, once an evaluation has been conducted, are set with key players excluded. The argument is that state policy on evaluation is not mediated,

adapted, implemented, adopted, or contested but mandated at moments and places that suit the government. Policy flow is therefore unidirectional and subject to a diverse range of influences and interests. This may mean that the gap between evaluation policy creation and implementation may be wide and growing. Due to lack of involvement, participants are disenfranchised in that they are not empowered to contribute to evaluation processes and procedures. This has effects on the Joint Committee of *standards* criteria which desire that all stakeholders be included at all levels of evaluation planning and operations.

The privileged position of the government does not always guarantee successful implementation of evaluation results. There are two explanations for this:

- The results have to be interpreted at various levels of administration, such as, K.I.E, permanent secretary in M.O.E, and the Minister of Education itself, and central the government.
- Interpretations are at different moments and places as the central authority is both a combination of competing ideas and implosions subject to diverse range of influences from the public, scholars and the donors

Promoting programme improvement and development is the principle aim of evaluation in programme evaluation process. It lies at the heart of the evaluation practice. It provides the framework in which programmes progress, may be charted and expressed. Considering the object for evaluation in this particular research, learning becomes the principle aim of education and any school activity. Evaluation lies at the centre of this process, providing a framework in which education objectives, specifically those related to curriculum content, may be set and learning progress chartered and expressed.



Mediating dialogue enhances evaluation professional skills and helps schools as a whole, to strengthen learning across the curriculum.

Looking at the evaluation arrangement, three agendas do arise. These are formative, summative and needs assessment, each showing different notions of accountability. All are closely associated with policy and the locale of the evaluand. This includes the school, with emphasis on reorganisation of education, the curriculum workload, and teacher professional potential, which is determined by the differing views related to government interests and programme aspects other than professional needs and functions of evaluation, thus the evaluation agenda Collective concern advocated by the Propriety\_Standards that address service delivery in terms of public good and consumer interest are not emphasised. Every time a need for evaluation surfaces, it means a series of disputes leading to conflicts as well as intense controversies. The results of the summative evaluation are most public and visible in this regard.

In these manifestations evaluation may be construed as a vehicle for disputes and legislating social order. From other perspectives, perpetuating heterogeneous systems of values has resulted in expressions of diversity spreading across the system defying encapsulation and so donor interference and public outcry.

As programmes get transformed from individual to national ones, in the current paradigm shift, the move towards mainstreaming evaluation due to donor demand for results for the purpose of results-based-management is becoming the focus for debate. The problem is

that much as outside influence through donor intervention could have far reaching influence in changing the way evaluation is practised, donor intervention is not systematic and it is guided by self interests that are sometimes not functional in evaluative terms. Unless this situation is critically analysed, this mode of intervention will not be beneficial to the development of evaluation practice and the status quo will be maintained for years to come unless evaluation principles are upheld by the donor.

There is also the need for an environment that facilitates and incorporates programme evaluation into government systems through good governance and a supporting infrastructure and appropriate technology. The relationship between donors and beneficiaries should be aligned and initiatives for evaluation capacity building made a priority. Beyond this there is need to make programme evaluation mandatory through legislation in all government core activities.

## **5.1.2 Summary of the Analysis on the Standards' Criteria** (Theoretical foundations of evaluation practice)

### **5.1.2.1 Utility\_ Standards**

It is generally stated that the purpose of undertaking evaluation is to help authorities make wise decisions. For an evaluation to be valid, the results must therefore be utilised. The *standards* are presented in order of importance with the *utility standards* ranking first. The standards focus on whether or not evaluation serves practical information needs and whether they guide evaluation so that it may be informative, timely and influential. The study revealed a situation whereby at the level of *utility* as prescribed by the

*standards*, the attribute that pertains to *use* was worse off in that most of the criteria were not met (out of 7 only 4 were *Partially Met*, while 3 were *Not Met*).

The findings from the study show a pattern of non-use of results, and very little activity is apparent after an evaluation is accomplished. This creates a problem in that stakeholders, that is, the civil society, the media and education scholars seem to be more interested in immediate and direct impact of the evaluation results. In fact, the literature on evaluations in developing countries confirms that evaluation results are generally not utilised in decision making (Asamoah, 1988).

A general consensus on evaluation use is that evaluations tend to have *incremental use*, that is, use implemented gradually over a period of time. In analysing the evaluation practice further and as the study shows, evaluation results are utilized, but, it can take up to ten years to implement important programme aspects. This may reflect a trend towards incremental use. Enlightenment, that is, many uses occurring in the long term to influence the design of programmes in future, does not seem to apply. The demand is however for *instrumental use*, that is, making direct and immediate changes to programmes. This creates a problem since the evaluation function is to determine the value of worth that involves making decisions on the way forward (as opposed to stagnation).

The Kenyan situation presents a unique situation regarding the theoretical issues arising from the literature. The situation is characterised by basically all the key factors such as not receiving optimum support to improve status and poor implementations of

programmes. The evaluand, that is, the 8-4-4 system of education is a case in point as reflected in the metaevaluation results. Furthermore, the fact, that evaluation is viewed more for its own system (power play) than scientific advantage. Certain elements are desirable when carrying out evaluations for use. These are identification of users before carrying out the evaluation; frequent contact with users; and participants both during planning the evaluation and while conducting the evaluation; providing interim results; translating findings into actions; disseminating results through various channels; and presenting data in policy debates. All these elements were not met.

The Kenyan evaluation practice presents a situation that is not simple, where political expediency takes precedence over the value of information, a situation that is more difficult to measure and demonstrate explicitly. The idea that most national programmes are controlled by the central government who decide on:

- who calls for evaluation
- who determines who should conduct evaluations
- who determines how evaluations are to be conducted
- who makes decisions on whether to utilise evaluation results, and
- who puts the evaluator in a challenging position.

The fact that evaluations are viewed with suspicion by the government affects the judgement, hence the quality of evaluation use. As a result, the government is put on the defensive. A lot depends on the outcome of the evaluation which in turn depends on whether the outcomes are in tune with the government's frame of reference. This creates the initial problem for utilisation of evaluation results.

Looking at the large evaluations in Kenya in the last ten years, one realises that evaluation results are available for years before political, economic and social conditions allow their use. Mention was made to the effect that over this period, evaluations have not been disseminated with exception of the latest needs assessment carried out to aid the Koech commission(2000) of inquiry. The issue of overloaded curriculum had been raised over time. However, it was not until ten years later that action was taken to reduce reasonably the subjects taught in schools. Even then it needed pressure from donors, civil society, and the opposition political parties to pressurise the central government to acknowledge need for change.

Accordingly, successful evaluators are those who have made clear to all concerned how evaluation is to be used and its level of application. This is not the case in Kenya. As a result, the mode in which the results are generally released presents a communication overload as the results are not released as desired, creating anxiety and suspicion. The short-term instrumental use, which culminates in direct and immediate changes, is more laudable as in most cases there is urgent need for immediate returns to which the uses are obligated.

This is not to say there is no place for enlightenment use, which is more appropriate in promoting academic theorisation. However, the evaluators are not professionally aware of the implication of differing levels of use and if evidenced, its application is more

incidental and not planned for. The Kenya evaluation practice presents a situation where there are still unresolved issues that need to be researched regarding *utility\_standard*.

In all this, credibility is important in order to meet the criteria for defending utility, practicality, integrity and technical adequacy. It is also important beyond credibility to be able to conceptualise evaluation principles and to translate these into evaluative principles. The implications of this was realised when analysing various *standards* that desire credibility for proper implementation and interpretation. These are inadequacies in U1 \_ *Stakeholder Identification*, F2 \_ *Political Viability*, A2 \_ *Context Analysis*, and A3 \_ *Described Purposes and Procedures*.

This discussion on evaluation practice in Kenya shows there is a problem with the *standards* addressing evaluation impact. While evaluation could be used to influence policy, the decision on policy does not (always) influence evaluation practice locally. Responsible policy makers need to proceed with implementing their results both at the policy and practice levels. The Kenyan evaluation practice reflects non-responsiveness of evaluation, a situation that depicts no influence in terms of decisions made and follow-up.

Evaluation theorists insist that the major pragmatic justification for evaluation is use without which evaluation loses the tie to policy that gave rise to the evaluation field.

### 5.1.2.2 *Feasibility\_Standards*

From the foregoing discussions, it shows that evaluation is not just a process of methodological and technical activity but also involves interpersonal sensitivity and political influence generally referred to as social ecology of evaluation context. The results show that a number of factors pervade every aspect of evaluation practice from initial decision to evaluate through data collection up to utilisation of evaluation results.

The results show that *F3\_Cost Effectiveness* was not carried out, and *F2\_Political Viability* and *F1\_Practical Procedures* were not met either. Numerous errors were noted, for example, not enough time was allocated to in carrying out evaluations, and stakeholder perspective was not taken into consideration, thus making the evaluations insensitive. The issue of non-utilisation of results has been raised. This makes one question the cost benefits when a task such as carrying out evaluation, that costs money, does not finally result in improving programmes and people's lives. Without a appropriate contractual agreement, those carrying out evaluations are not made accountable, posing a danger in possible evaluator bias.

Certain biases arose and are found in the *standards*. These were:

- what the standards criteria refer to as giving appearance where the evaluation is biased towards one stakeholder.
- failure to take care of both formal and informal organisational power structure.
- and completely ignoring cost effectiveness.

According to the *standards*, costs refer to social value and monitoring of all human and physical resources used to carry out evaluations. Benefits on the other hand denote the value of all the results derived from evaluations which involve identifying effective and ineffective programmes; fostering understanding of activities and how they are perceived. However, authorities in the field acknowledge that cost effective analysis is difficult to manage as outcomes of evaluations in this particular context are numerous, intangible, and often valued differently by different stakeholders.

What is emerging are factors of inadequacies related to U1\_ *Stakeholder Identification* and A2\_ *Context Analysis* that are important in maintaining a rapport between the various perspectives and processes. This has a direct bearing on the nature and manner of communication practices. For example, it is important to know that sanctions and cooperation need to be obtained from those in power; reporting style must be clear to everyone from the outset; client(the government in this case) should agree on ways to maintain open disclosure and modes of resolving conflict should they arise. The evaluation as practised in Kenya does not take into consideration these elements. None of these were put in place and stakeholders were not involved confirming poor communication practices. The study exhibits many examples of poor communication practices with no incorporation of appropriate practices. Consequently, the evaluators are caught up in technical matters other than communication, thus leading to harmful evaluations.



Problems were experienced in a number of *standards* which may be due to a bias at some point in the evaluation process. This could have a direct bearing on influencing evaluators practices and decisions. Some of these are: F2\_ *Political Viability*; P5\_ *Complete and Fair Assessment*; P6\_ *Disclosure of Findings*; P7\_ *Conflict of Interest*. These *standards* show that evaluators in Kenya generally conform to the social context other than the payment they get for carrying out evaluations. This situation is referred to as co-optation, a practice that is found in externally conducted evaluations. This may raise questions in regard to the independence of the evaluator. The researcher did not find evidence of such conflict.

It is important to understand the potential bias resulting from the context in which evaluation is being conducted so as to be prepared to address the problems related to *feasibility standards*.

### 5.1.2.3 *Propriety\_ Standards*

The *propriety standards* are placed third in their level of importance and are not synonymous with ethical codes of concerns. The other elements, that is, utility, feasibility and accuracy relates indirectly to ethical principles. The *standards* as described are *guidelines* rather than *principles* of ethical practice and relate to privacy, protection of subjects and freedom of information. There were problems with the key *standards*, such as, *Propriety\_standards* none of which were fully met such as P1\_ *Service Orientation*, P5\_ *Complete and Fair Assessment*, P6\_ *Disclosure of Findings*. *Standards* related to

P2\_*Formal Agreements*, P3\_*Rights of Human Subjects* and P4\_*Human Interactions* were only partially met.

There are evaluation theoretical issues that have been identified in evaluation literature that may influence the application of ethical standards that may have a bearing on application of the *standards* to the Kenyan situation. Which could put reservations to their practicability. According to House(1993), ethics best applies to specific and concrete situations because a thousand factors can make a difference in particular situations and that ethical concerns cannot be exhausted. Making decisions on one ethical issue involves trade-offs about which few authorities agree on. Brown and Newman (1992) state that there has not been extensive literature regarding ethical issues in evaluation and that what exists calls for dialogue and discussion to further guide ethical practice in evaluation. This situation is further compounded by the fact that within the African context resources are scarce and experts in evaluation are few. The Kenyan situation therefore poses a challenge in application of this *standards* due to the typical African context where variations where variations are not only politically felt but also regionally (consider urban/ rural) and individually in terms of differences amongst students within the same class.

The application of the *standard* is further complicated by the theories on which the ethical *standards* are based which reflect the American context. Shardish et. al. (1991) state that the ethical *standards* as prescribed are based on theory-related terms, such as, producing the greatest good for the greatest number, obligation, acting properly or justly,

justice that emphasise liberty or freedom and autonomy. According to Taut (2000), the notion of propriety (or morality) relies on *standards* that are equally applicable to everybody, a position typical of the North American values on which the principles of propriety *standards* are based. It would seem that hierarchical societies on which LICs base their values ultimately present violations of the *standards*. Accordingly, these systems represent the legitimacy of inequality of individuals and groups that stress situation, or context information when judging behaviour.

Taut (2000) states that, in calling for serving all stakeholders, it would seem that the range of service delivery is out of reach of the evaluator in the Kenyan context. This results in the rights of the individual being considered subordinate to the greater good of the group. Also, the codes of *directness*, *openness* and *completeness* may not be applicable in such hierarchical structures. The author further states that even some countries that would be expected to have a more open society such as *Australia* and *South Korea* do not always take these principles as their first concern, reflecting limitations as to the extent to which *directness*, *openness* and *completeness* can be applied internationally. However, this does not eliminate the fact that within the *standards* criteria, the Kenyan evaluation practice and operations do not meet the requirements of propriety *standards*.

As to dealing with conflicts openly and honestly, dire consequences have been taken whereby evaluators have on occasions been victimised in the LICs' context (Asamoah, 1988). The Joint Committee makes an exception of this *standard* and states that, where

disclosure would endanger public safety or bridge individual freedom, the information could be withheld. This arises from the acknowledgement that evaluations have a potential for conflict and therefore, the problem is not how to avoid conflict but how to deal with it. The results show that the evaluation reports may have avoided addressing conflicts for these reasons. It could therefore be concluded that those conducting evaluations in Kenya are within the limits of the *standards* criteria in taking the position they have. However, the debate continues and theories that address these issues should be researched as they cannot be discounted from the function of evaluation.

What is emerging from this discussion is that propriety *standards* which were generally not met, are difficult to apply to hierarchical cultures such as Kenya, in fact, it may not be possible. The extent to which this is detrimental to evaluation practised in such contexts is the question to ask.

In this study, the researcher set out to determine the appropriateness of evaluation practices in Kenya by looking into evaluation practices in education using the *standards* as the criteria for measurement. Certain limitations were recognized that do not imply inappropriate practices but were determined by the situation at hand. However, by not collecting appropriate data either due to negligence or as a result of restriction arising from the context in which evaluation was practised, objectivity of the results was compromised. Also, providing only one interpretation because since stakeholders' views are not generally upheld, the autonomy of the client (or the central government) was upheld.

This created a problem in applying the *standards* as they require information regarding impact to be balanced between the stakeholders and the client, objectivity of the findings, as well as full disclosure of findings. To this end, the question to ask is, “to what extent was this detrimental to evaluation practice?” The answer is that it thus represents inadequate coverage by the *standards*.

#### 5.1.2.4 *Accuracy\_ Standards*

Except for certain explicit elements, it was difficult to interpret the extent to which the *standard* was violated given the practical difficulty involving matching good evaluation designs to practical circumstances that are manifold in LICs. The results of the research show that out of the twelve *standards*, A3\_ *Described Purposes and Procedures*; A6\_ *Reliable Information*; A7\_ *Systematic Information*; A-10\_ *Justified Conclusions*; A\_ 11 *Impartial Reporting* were not met while the others were only partially met there being no evidence of metaevaluation.

Any set of *standards* generally assumes that it is possible to distinguish good and bad technical aspects of evaluation. In considering purely technical decisions such as determining the size of the data, there are no easy answers. This is further complicated by the evaluation data that depend on knowing the nature of the information that is not readily established and impossible to know. Looking at the issues related to validity, one wonders how much should be traded. It was not easy to determine whether it would have been appropriate to use a randomised design instead of a representative sample. Because

of variability, the Kenyan context is not appropriate for randomisation. The choice between *homogeneity* and *innovation* is very real. Heterogeneity results in varied situations generally not found in ICs in which most of the literature on evaluation methodology is derived.

Time and resources also prohibit widening the sampling net which would ensure similarities of variables to be analyse. However, in education the context is not so critical as there are advantages in terms of determining sampling and availability of respondents due to the nature of the school setting (vs. community setting) that can be better determined.

The ensuing discussion may explain why a number of standards were met as the trade-offs. However there is no guideline in the evaluation literature giving guidance on how much trade-off should be allowed. The *standards* are also silent on this issue except for some explicit elements.

Contextual factors play a great role in conceptualising evaluation. For example, overworked and underpaid teachers result in apathy. This situation as established from the evaluation documents, characterises teachers in Kenya. This leads to evaluation measures being kept at a minimal, forcing questionnaires to be short and non-complex or else teachers would not complete them properly or even attempt to fill them. The interviews revealed that one of the reasons for short questionnaires was precisely that. However, the questionnaires should have covered more issues so as to contribute to better

knowledge of the subject at hand. Lacking in the questionnaires were also issues related to special and time consuming processes such as classroom interaction processes.

The unanticipated frequent circumstances also play a role, for example, in cases where children are sent away for school fees, or are involved in sports. Even then, success in one situation may not necessarily mean success in another because of the stated variations. Evaluators often overlooked the variations that existed yet this can affect the result. The evaluators did take certain measures into consideration, such as, those related to disruption of classes during examinations. Field work was not carried out during such times.

Looking at the way evaluations were carried out, the practice was conventional to the extent that the designs were based on the requirements of the objectives of the evaluation. The objectives did not vary much across the different evaluations. There was a lot of reliance on precedence in terms of what was done previously which is not necessarily bad but can be a problem as it has been agreed that variations outcomes and processes depend on the context and purpose in which evaluation is conducted. This is in relation to the period one conducts evaluation matters. By relying on precedence one could miss important considerations necessary for responsiveness of an evaluation.

All evaluations carried out to determine the appropriateness of 8-4-4 system in Kenya resulted in a recommendation to reduce the work-load as a result of a loaded curriculum. This was not good for the political party in power which was determined to keep the

system as it was. Many such cases do arise from evaluations of government programmes in LICs.

Methodological flaw was also evident due to demand for rapid evaluation results. This hinders the *slow, deliberate, integrated* process advocated by the *standards*. One therefore found a methodological approach that was governed more by *expediency* in terms of how suitable other than *scientifically* appropriate it was.. This has resulted in a more microcosmic(minute) versus macrocosmic (wholesome) approaches to evaluations.

The most affected were the externally conducted evaluations. As far back as 1976, Brickell, H. M., realised that political factors did influence methodology in terms of the sampled populations, the nature and amount of data to collect, the designs, the interpretations, the recommendations and even the wording of the reports. This study confirms Brickells observations and has been illustrated at various points.

The other important process in evaluation is validity and reliability. Evaluators are consistently exposed to questions of value which influence their purpose and design of studies, choice of methods and use of results. The value option chosen in evidencing validity and reliability were not consistent with the *standards* procedure. This meant that evaluations did not have reliability checks, thus diluting their credibility. There was no evidence of reliability measures in the evaluation documents. The evaluations involved administering tests to determine student achievement. The results show the use of pilot testing but no indication of reliability and validity applications.



The results also reflect use of simple calculation approaches such as means, and standard deviations regarding quantitative information and percentages regarding qualitative information. This raises the question as to whether the unit used were the most appropriate (or the easiest to handle). In most cases, the school was the unit for selecting comparison groups, the classroom the unit of programme implementation, and the student the unit of analysis. Regional disparity in terms of urban/ rural spread, facilities both physical and financial, were considered. However, special units such as systematic analysis of gender issues related to aggregated data, curriculum content review and analysis, as well as classroom learning observations were not taken into account. Since the issue of non-competence of evaluators was ruled out, it was be assumed that negligence could have been the problem. The results of the evaluations are therefore, incomplete in terms of impact of intervention.

In view of all this, the credibility of the evaluator becomes important. It was established that all the evaluators had research background with at least a course in evaluation with strong research related degrees. However, going through the studies, there were shortcomings on the evaluators conduct in that some cases reflected negligence as stated earlier, for example, because of their research background, the evaluators tended to preserve in evaluations technical qualities that make good research studies. It should be noted that there are other qualities such as political expediency, psychometric measurement issues involving tests, content reviews and micro elements of the evaluand. A case in point is classroom learning observations, an analysis that was avoided. It is important for evaluators to have knowledge of how various values influence design and

conduct of evaluation studies and the results. Of significance was lack of many desirable qualities related to special principles of evaluation methodology. Consider issues of *Utility\_ Standards* whereby responsiveness of evaluations were not attained as the statements made in terms of outcome and recommendations were neither specifically (as opposed to generally) stated nor tentatively (as opposed to speculatively) stated, thus reducing responsiveness of evaluations.

The above deliberations have a direct bearing on *standards* related to *justification* of conclusions. Issues of justification and therefore defensibility of results was not possible as information collected lacked comprehensibility, sound analysis and statistical logic desired by the *standards*. Much as the conclusions respond to the evaluation questions, the objectives did not address objectives related to overall education goals comprehensively or the general trends in international curriculum education needs such as gender issues. There were no conflicting conclusions and recommendations reported, nor negative findings made explicit, reflecting an element of caution in the interpretation and presentation of the results.

The important of stakeholder involvement is very strongly guarded in the *standards* criteria. There was little involvement of participation of stakeholders with no option for feedback at all critical levels of the evaluation process. Participants were included only in their capacity as respondents. This did not meet the criteria of impartial reporting.

Lastly, without reviewing the evaluation as this study has done, one would not be able to pin-point with such accuracy ensuing problems as K.I.E attempts to manage evaluation almost single-handedly. As noted, *metaevaluation* was not an activity of the evaluation process. This denied the evaluation process self- evaluation, a useful procedure in improving evaluation practice.

On the whole, the *Accuracy Standard* was better addressed than all the other standards. However, there were still problems which are necessary in meeting the *standards* criteria. One advantage the evaluators had was their background in and experience of social research applications that helped guide the technical aspect of evaluation practice.

## **5.2 Synopsis of the Metaevaluation Outcome**

The unique contribution of this research was to determine evaluation practice in complex applications within context of LICs and theoretical approaches to evaluation practice and process against the Joint Committee of *Standards*(1994). This was in regard to Kenyan context and the challenges the context poses. The results show that, there is need to appreciate and understand the constraints and implications of issues that plague evaluation practice. The process of *meatevaluation* has brought to light clinical issues that need to be addressed for evaluation to meet its function in ensuring effectiveness of programmes.

The results show that, the notion of *accountability* was not easy to come to terms with for a number of reasons:

- The framework that arises as evaluators carry out their work
- The notions of power and control which leads to constraints and sanctions
- The means to achieve the desired end that required trade-offs
- The notion of what the end should be in terms of programme needs, technical adequacy, evaluation methodological needs whereby evaluation methodological needs were not met resulting in a non-responsive evaluation.

The outcomes of the *metaevaluation* show that the approaches that reflect programme needs and technical adequacy based on social science inquiry were addressed but not the evaluation principles and methods of service delivery and utility. Therefore the programme worth determined using differing criteria which is also based on interest of different groups and different ways in which one player exercised control over the other. Diversity and complexity are at odds with the concept of the *standard* as a tool for legitimising evaluation in terms of agreed measures or norms, an important and continuing feature.

Evaluation also exists within social and psychological concerns whereby, there is an association with *high standards* and with *standard bearers* around whom members of the various groups rally, more so in times of conflict. This is a continuing feature that characterises programme evaluation practice in Kenya. The evaluation process may probably exert a more permanent influence if it were not for control by the government.

The question as to whom accountability should be attributed did come up. Should it be to the broader group, and if so, would this be possible in LIC context. The *standards*

advocates the broader society. The extent to which an evaluator is responsible to a spectrum of stakeholders therefore becomes a moral issue. This has a bearing on the extent to which disclosure of findings, dissemination, release of results to the general public and interested parties would be practically possible. Immediately this issue is addressed, it takes those responsible for evaluation from the arena of responsibility into the arena of relationships between different interest groups such as evaluators themselves, the client, and the wider profession. The fact that the government is the overall authority creates a problem and one wonders if it is possible in such a context to serve the broader group. There is also the concern of what this means in practice as the evaluation must conform to its initial function where the project delivers data which is useful and has potential to advance understanding the conduct of evaluation. There are therefore, conflicts in a number of ways conditions.

In the event that it is not always possible to apply evaluation principles in all contexts, more so in LICs' world contexts, there is a need to review the evaluation practice and possibly recommend ways in which one can carry out evaluation and still address the difficulties. With some theorists advocating re-conceptualisation of or realistic approach to evaluation, the researcher agrees with Cronbach(1982) who states that the evaluators aim should not be to diminish or control the effects in this context, but to use the in-context information to illuminate the evaluation so that ultimately, the programme can be improved. Evaluation has always had to deal with a number of constraints that have affected its practice. According to Mowbury et. al. (1998), there are no specific answers

to ensuing problems arising or particular approaches to guide practice. The *standards* is also not explicit on how to handle these difficulties.

The rationale for doing applied work is to influence actions in terms of decision making by using the findings and conclusions provided by evaluation. In designing and conducting evaluation, one needs not only a *accuracy\_standards* or technical inquiry skills but also familiarity with the *utility\_standards*, *propriety\_standards* and *feasibility\_standards*.

The results show that no method can guarantee the acceptance or eventual use of empirical data, and that, in the Kenyan context, results of the research show lack of many desirable qualities related to special principles of evaluation process. Notable ones and basic to programme evaluation are the *utility standards*. The other two *standards*, that is, *feasibility standards* and *propriety standards* were also flawed whereby certain basic elements desirable in any evaluation such as the social perspective of evaluation context, communication style, sensitivity to political viability, determination of cost effectiveness, self-evaluation and disclosure of findings were not met. The researcher makes the conclusion that the evaluation practice as desired by the Joint Committee does not meet the set criteria as most standards were only partially met or not met.

However, to be explicit, one would have to re-evaluate and follow closely the discussions to establish the fine grain conflicts arising and the positive elements identified in the study to be explicit. Despite problems in application of *propriety standards* which are

based mainly on North American values, and the characteristics of evaluation context in Kenya, the *standards* tend to enumerate various technicalities worth investigating and bringing to light in order to further shed light on some of the grey areas.

The impact of intervention leads to possible inaccurate and unacceptable conclusions as information on which the results are based lacked comprehensibility and sound analysis and statistical logic desired by the *standards*.

### 5.3 Conclusion

The conclusion is made to the extent that programme evaluation practice in Kenya does not adequately meet the requirements of acceptable evaluation practice as prescribed by evaluation theory and literature, and that, the context within which evaluation takes place does not facilitate the process of evaluation. The overall judgement the results of the evaluation practice in Kenya is flawed is a fair one from the point of view of the *metaevaluation* of this study. Not meeting the *standards* criteria adequately implies that the evaluations are flawed to the extent to which the *standards* were met. The issues that arose are listed as follows:

- Evaluation operations and services are controlled and sanctioned by issues of power which are directional and subject to diverse range of influences and interests.
- Evaluation is determined by government interests and programme aspects other than professional needs and functions of evaluation, such as monitoring, formative and summative evaluation.

- A hierarchical society on which Kenya bases its values ultimately present violations of the standards, thus putting a range of service delivery out of reach, such that, serving all stakeholders, code of directness, openness and completeness in *utilising* results are not taken as the first concern.
- Evaluation practice loses tie to policy that gave rise to the evaluation field as results are not *utilised* to the extent that they are not informative, timely or influential; instead political expediency takes precedence over value of information
- Evaluation theoretical issues have a bearing on the application of the *standards* putting reservations on their *practicality*, for example, the propriety\_ *standard* which is affected by the varied contexts relies on trade-offs since ethical concerns cannot be exhausted in a situation where politics governs practice, where resources and experts are scarce and few.
- Evaluators are not professionally aware of a number of implications that is why evaluation is guided by social research approaches and expertise and not evaluation methodological needs, thus reflecting non-responsiveness of evaluation, a situation that depicts no influence in terms of decisions made and follow-up
- Methodological flaws arise as they are influenced by need for rapid evaluations, resulting in an approach governed more by expediency in terms of how suitable than how scientifically appropriate the evaluation approach is. This leads to value options chosen, such as validity and reliability, not being consistent with the *standards* procedure and evaluation literature.
- The methodological approach reflects technical sense of evaluation, with attributes of measurement and not impact, in managerial system that is disempowering as the manager has powers over evaluation process and utility of results. This does not reflect collective responsibility but an individual one.

However, it would appear that the findings are not always because of inappropriate practices but are determined by the situation at hand. The Kenyan evaluation practice



presents a situation that is not simple with unresolved issues that need to be regarded as possibly new forms of evaluation approaches in a unique context. The range of practice needed to meet the *standards* criteria seems to be out of reach of the evaluator as theoretical evaluation issues that influence practice have not been addressed in the evaluation literature to guide practice. The *standards* are also silent on these issues.

Findings regarding Kenyan evaluation practices are corroborated in earlier writings by:

- Scriven(1994), that, technical sense of evaluation persists to day.
- Patton(1985), that, there is no one way to do evaluations and that, evaluators have to deal with complexities arising in terms of management, policy, value economics and psychological questions.
- Smith(1991), that, issues of whose values and criteria to choose from are still noticeably absent in the literature to guide practice.
- Asamoah(1988), that, political expediency strongly influence decisions about policy and planning.
- Asamoah(1988), that, results having distinct implications policy, are not always put to use
- Asamoah(1988), that, evaluations mainly serve donor interests and rarely do they attend to national interests, thus questioning the practical utility of impact evaluations in LICs.
- Ginsberg(1988b), that, evaluators are put in a context of completing values related to special interest groups which may reflect differences in epistemology and philosophy.

## 5.4 Recommendations

The following are the recommendations emanating from the study.

- **Evaluation context**
  - ✓ There is need to develop and *legislate* a policy on evaluation.
  - ✓ There is need for collective concern to address issues related to *service delivery* in evaluation that include evaluation functions, such as monitory, formative and summative functions.
  - ✓ There is need to create an environment that facilitates and incorporates professional needs of evaluation and its appreciation in order to reduce conflict and controversies that arise when evaluation is called for.
  
- **Theoretical basis of evaluation practice**
  - **Utility \_ *Standards***
    - ✓ Patterns of *evaluation use* should be studied so that results serve practical information needs that guide evaluation to make them informative, timely and influential.
    - ✓ There is need to design strategies for *disseminating* results through various channels of programme managers, the government and the civil society to strengthen utility of evaluation results.
    - ✓ Strategies should be devised to make it possible and influence Policy makers to implement evaluation results both at the policy and practice levels so as to reflect responsiveness of evaluation.

### □ Feasibility \_Standards

- ✓ Elements that hamper *communication practices*, such as interpersonal sensitivity and political influence of evaluation context should be taken into consideration by maintaining a rapport between various perspectives, processes and stakeholders.
- ✓ *Reporting style* should be agreed upon from the outset; and ways to maintain open disclosure and modes of resolving conflict agreed upon.
- ✓ Those managing evaluations should be made *accountable* through desirable contractual agreements in regard to cost effectiveness, that is, social and monetary value of human and physical resources.
- ✓ *Stakeholders* should be involved throughout the evaluation process.

### □ Propriety \_Standards

- ✓ There is need for dialogue and discussions to guide *ethical practice* within the Kenya context due to inadequate guidance on application of the *standard* in order to establish the extent to which the *standard* can be applied within a hierarchical system of governance, such as that of Kenya, a system which stands for the legitimacy of equality of participation of individuals, groups and institutions resulting in empowerment.
- ✓ Theories that address *propriety needs* and the extent of their applicability to evaluation practice in Kenya should be researched more so as to relate to disclosure of findings.

□ **Accuracy \_Standards**

- ✓ There is need to mandate knowledge of evaluation *methodological approaches* beyond the social research techniques of evaluation process by those carrying out evaluations as well as those implementing programmes.
- ✓ Evaluations should be *innovative* and not merely conventional through mere application of objective identification of evaluation approaches. This would address the variations arising in programme systems and operations in terms of process, outcome, and impact that characterise the Kenyan context to ensure responsiveness of evaluation and justification of conclusions.

□ **Metaevaluation**

- ✓ Evaluation mechanisms should be mainstreamed in all functions of government and self-assessment, making *metaevaluation* mandatory.
- ✓ There should be a deliberate plan to develop appreciation of evaluation practice in itself in terms of its professional functions and how it can contribute to programme development and reform.
- ✓ There should be capacity building for all with a stake in evaluation, hence *metaevaluation* process right from programme managers and dispensers, the civil society and the client.
- ✓ There is need to carry out further research into micro-aspects of issues that plague evaluation practice in view of the position taken that; and the challenging context in

which evaluation operates in Kenya in order to create more insight and to contribute to cross-cultural evaluation literature.

## REFERENCES

- Alkin, M.C. (1969). Evaluation theory development. *Evaluation Comment*, 2, 2-7.
- Asamoah, Y. (1988). Program evaluation in Africa. *Evaluation and Program Planning*, 11, 69-177.
- Balthasar, A, & Rieder, S. (2000). *Learning from Evaluations*. *Evaluation*, 6(3), 245-260.
- Barnhart, T. (1993). The World Book Dictionary. Chicago: World Book.
- Brass W. & Jolly C.L. (1993). Population dynamics of Kenya. Washington D.C.: Academy press:
- Bootzin, R.R., Bower G.H., Crocher, J. & Hall, E. (1992). Psychology today: An introduction. London: Mcgraw-Hill.
- Boruch, R. C., & Cordroy, D. S. (1980). An appraisal of education program evaluations: Federal state and local agencies. In Asamoah, Y., (1988). Program evaluation in Africa. *Evaluation and Program Planning*, 11, 169-177.
- Brown, R. D. & Newman, D. L. (1992). Ethical Principles and Evaluation Standards: Do They Match? *Evaluation Review*. 16(6), 650-663.

Chapman, D. W. & Boothroyd, R. A. (1988). Evaluation Dilemmas: conducting evaluation studies in developing countries. *Evaluation and Program Planning*. 11, 37-42.

Child F.C., (1977). Small-scale rural industry in Kenya. African studies center, University of California: Los Angeles.

Clemson, D.W. (1985). Supporting evaluation in Kenya: Reflections on experiences with the Kenya adult literacy programme. *International Journal of Education Development*, 5(3), 245-248.

Conner, R.F. (1985). International and domestic evaluation: Comparisons and insights. In Patton, M.Q. (Ed.). Culture and evaluation. *New Directions for Programme Evaluation*, 25, 19-28.

Cook, T.D., (1987). Postpositivist critical multiplism. *Evaluation studies: Review Annual*. 12. (Ed). Shardish, Jr., W.R. & Reichardt, C.S. London: Sage.

Cook T. D., Shardish, W.R., Jr. & Reichardt, C.S. (1987). On what scientists know and how they know it. *Evaluation Studies*. 12, 378-393.

Cook T. D. & Gruder. (1978). In Asamoah, Y. (1988). Program evaluation in Africa. Evaluation and Program evaluation in Africa. *Evaluation and Program Planning*. 11, 169-179.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.

Cronbach, L. J. (1963). Designing evaluation of educational and social programmes. San Francisco: Jossey-Bass.

Cronbach, L. J. (1982). Designing evaluation of educational and social programs. San Francisco: Jossey-Bass.

Cronbach, L.J., Ambron, S. R., Dornsbusch, S. M., Hess, R. D., Hornik, R.C.,

Phillips, D. C., Walker, S. S. (1980). Toward reform of program evaluation.

San Francisco: Jossey-Bass.

Curthbert, (1985). Evaluation encounters in third world settings. A Caribbean perspective. In

Eastmond, N. (1991). Education evaluation. The future. *Theory and Practice*. 30(1), 74-79.

Daily Nation News Paper. (2000, March 27<sup>th</sup>). On 8-4-4 system of education.

Dooley, D., (1984). Social research methods. Prentice-Hall : New Jersey.

Eastmond, N. (1991). Education evaluation: The Future. *Theory and Practice*, 30(1), 74-79



Ewusi, (1978). *Evaluation practices of education programmes reviewed by the joint dissemination panel of review 1980-1983.*

Fettermen, D. M. (1994). Empowerment evaluation. *Evaluation practice*. 15(1), 1-15.

Forss, K. & Carlsson, J. (1997). The quest for quality\_or can evaluation findings be trusted. Special Issue: European Evaluation Society Conference. *Evaluation*. 3(4), 481-502.

Ginsberg, P.E., (1988). Evaluation in cross-cultural perspective. *Evaluation and Program Planning*. 11, 189-195.

Glaser, L. & Strauss, M. (1985). *Qualitative research methods*. London: Sage.

Government of Kenya, (2002). National development plan 2002-2008. Effective management for sustainable economic growth and poverty reduction. Nairobi: Government Printers.

Guba, E. G., (1969). The failure of education evaluation. *Education Technology*, 9, 29-38.

Guba, E. G., (1981). Investigating reporting. In N. L. Smith (Ed), *Metaphores for evaluation: Sources of new methods. New perspectives in Evaluation* (Vol.1). Beverly Hills, CA: Sage.

Guba, E. G., & Lincoln, Y.S. (1981). Effective Evaluation. San Francisco: Jossey-Bass Publishers.

Guba, E. G., & Lincoln, Y.S. (1989). Fourth generation evaluation. London: Sage.

Herman, J., Morris, L. L. & Fitz-Gibbon (1987). Evaluator's Handbook. London: Sage.

House, E. R. (1980). Evaluating with Validity. London: Sage.

House, E.R. (1993). Professional evaluation: Social impact and political consequences.

London: Sage.

Huey-Tsyh, C. & Rossi, P. H. (1987). The Multi-Goal, Theory Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science. *Evaluation Studies: Review journal*. 12, 38-53.

Jang, S. (2000). The Appropriateness of Joint Committee Standards in Non-Western Settings. In *The Programme Evaluation Standards in International Settings*. Ed. By Russon, C. The Center for Occasional Papers Series.

Joint committee on standards for education evaluation. (1994). 2<sup>nd</sup>. Ed. The Programme Evaluation Standards: How to Assess Evaluation of Education of Education Programmes.

2<sup>nd</sup>. Ed. London: Sage.

Kelly, K. & Vlaenderen, V. (1995). Evaluating participation processes in community development. *Evaluation and Program planning*. 18(4) 371-383.

Lewy, A. (1981). The Scope of Education evaluation: An Introduction.

New York: Breach, Science.

Lincoln, Y.S. & Guba, E.G., (1987). Trustworthiness and authenticity in naturalistic evaluation.

*Evaluation studies: Review annual*. 12. London: Sage.

Linn, R.L., (1978). Demands, cautions, and suggestions for setting standards. *Journal of*

*Education Measurement*, Vol: 15(4).

Lipsey, M.W., Crosse, S., Dunkle, J., Pollard, J., & Stobart, G., (1985). Evaluation: The state

of the art and the sorry state of science. *New Directions for Program Evaluation*.

27, 7-28.

Lynch, K.B. (1988). Evaluation practices of education programs reviewed by the joint

dissemination review panel, 1980-1983. *Evaluation Review*, 12(3), 253-275.

Madaus, G. F., Stuffelbeam, D. & Scriven, M. S. (1983). Programme Evaluation: A Historical

Overview. In Madaus, G. F., Scriven, M. S. & Stuffelbeam, D. L. *Evaluation Models:*

*Viewpoints on Education and Human Services Evaluation*. Boston: Kluwer-Nijhoff Publishing.

Marklund, S. (1984). Applicability of standards for evaluation of educational programmes,

projects, and materials in an international setting. *Evaluation and Program Planning*. 7, 355-362.

McCormick, R., & James, M. (1990). Curriculum evaluation in schools. 2<sup>nd</sup>. Ed. London.

Merryfield M.M., (1985). The challenge of cross-cultural evaluation: Some views from the field. *New Directions for Program Evaluation*, No:25, pp. 3-17.

Mertens, D. M. (1994). Training Evaluators: Unique Skills and Knowledge. *New Directions for Programme Evaluation*. 62, 17-27.

Ministry of Education (1964). Kenya Education Commission Report. Part 1: Ominde Commission. Nairobi: Government Printers

Ministry of Education (1976). Kenya Education Commission Report, Part II: Gachathi Report. Nairobi: Government Printers.

Ministry of Education. (1988). Report of the presidential working party on education and manpower training for the next decade and beyond: Kamunge report. Nairobi: Government Printers.

Ministry of Education (MOE). (1987). Education In Kenya: Information Handbook. Nairobi: Jomo Kenyatta Foundation.

Ministry of Education. (1988). Report of education administration conference. Nairobi: Jomo Kenyatta Foundation.

- Ministry of Education. (1999). Report of Commission of inquiry into education system of Kenya.  
Total Integrated Quality Education and training (TIQET). The Koech Report.  
Nairobi: Government Printers.
- Mowbray, C.T., Bybee, D., Collins, M.E., Levine, P. (1998). Optimizing evaluation quality and utility under resource constraints. *Evaluation and Programme Planning*, 21, 59-71.
- Mulusa, T. (1988). Evaluating education and community development programmes. College of Adult and Distance Education \_ Nairobi: University of Nairobi Press.
- Nevo, D. (1983). The conceptualisation of educational evaluation: An analytical review of the literature. *Review of Educational action through new orientations to evaluation. Research Studies* 53(1), 117-128.
- Nevo, D. (1984). Applying the evaluation standards in a different social context. *Evaluation and programme planning*. 7, 383-386.
- Newman, D.L. & Brown, R.D., (1992). Evaluations of evaluation standards: Frequency and seriousness of occurrence. *Evaluation Review*, Vol: 16, 219-234.
- Nielson, K. (1993). The confirmation of critical theory: The role of reflectiveness. *Human Studies*. Vol:16. pp. 381-397.

- Nilsson, N. & Hogben, D. (1983). Metaevaluation. *New Directions for Programme Evaluation*. 19, 83-95.
- Owen, J.M. (1989). Evaluation abstracts. Encouraging actions through new orientations to evaluation. *Studies in Educational Evaluation*, 15, 141-150.
- Patton, M. Q., (1985). Culture and evaluation. *New Directions for Programme Evaluation*. 25.
- Patton, M.Q. (1986). Utilised focused evaluation. 2nd. Ed. London: Sage.
- Patton, M.Q., (1988). Politics and evaluation. *Evaluation practice*, 9, 89-94.
- Patton, M. Q., (1997). Qualitative evaluation research methods. Sage.
- Payne, D. A. (1994). Designing Educational Projects and Programme Evaluations: A Practical Overview based on Research and Experience. London: Kluwer Academic Publishers
- Phillips, D.C., (1987). Philosophy, science and social inquiry: Contemporary
- Smith, N. L., Chircop, S, & Mukherjee, P. (1993). Considerations on the development of culturally relevant evaluation standards. *Studies in Educational Evaluation*. 19, 3-13.
- Popham, W.J. (1993). Education evaluation. 3rd. Ed. London: Allyn & Bacon.

Provus, M.M., (1971). Discrepancy evaluation. Berkeley, CA: McCuthan.

Sadler, D. R. (1985). The Origins and Function of Evaluation Criteria. *Education Theory*. 35(3), 285-297.

Scriven, M. (1967). The methodology of evaluation. In Stake, R.E., (Ed), Curriculum evaluation. *American education research association monograph series on evaluation*, Chicago: Rand McNally.

Scriven, M. (1975). Evaluation bias and it's control. Unpublished manuscript.

Scriven, M. (1978). *Goal-free evaluation in practice*. Paper presented at the annual meeting of the American Education Research Association. Toronto, Ontario, Canada.

Scriven, M. (1983). Evaluation Models: View points on educational and human services evaluation. Edited by Madaus, G.F., Scriven, M.S. & Stufflebeam, D.L. Boston: Kluwer-Nijhoff Publishing.

Scriven, M. (1991). *Evaluation Thesaurus*: 4<sup>th</sup>. Ed. London: Sage.

Scriven, M. (1994). Evaluation as a discipline. *Studies in Education Evaluation*. 20, 147-166.

Scriven, M. & Stufflebeam, D.L. (1981). (Eds). Evaluation models: View points in educational and human services evaluation. Boston: Kluwer-Nijhoff.

Sechrest, L., West, G.W., Phillips, M.A., Redner, R. & Yeaton, W. (1979). *Some neglected problems in evaluation research: Strengths and integrity of treatments.* San Francisco: Jossey-Bass.

Shardish, W. R. (1987). Sources of Evaluation Practice: Needs, Purposes, Questions and Technology. *Evaluation Studies: Review Annual.* 12, 521-553

Shardish, W. R. & Reinhardt, C. S. (1987). Programme Evaluation: The Worldly Science. *Evaluation Studies.* 12, 31-69.

Shardish, W. R., Cook, T.D., Leviton, L.C. (1991). *Evaluations of Program Evaluation: Theories and practice.* London: Sage.

Sharps, D.K. (1988). Curriculum traditions and practices. Worcester: Billings and Sons Ltd.

Smith, N.L. (1980). Bibliography on evaluation utilization. Portland, OR: Northwest Region.

Smith, N. L. (1991). The context of investigations in cross-cultural evaluations. *Studies in education evaluation.* 17(1), 2-21.



Smith, N.L. (1993). Considerations on the development of culturally relevant evaluations standards. *Studies in Education Evaluation*, 19, 3-13.

Smith, N.L., Chircopt, S. & Murkherjee, P., (1993). Considerations on the development of culturally relevant evaluation standards. *Studies in education evaluation*.19, 3-13.

Stake, R.E. (1967). The countenance of education evaluation. *Teachers College Record*. 68, 523-540.

Stuffelbeam, D.L. (1968). *Evaluation as enlightenment for decision making*. Columbus, OH: University of Nebraska Press.

Stuffelbeam, D. L., (1981). *A review of progress in educational evaluation*. Paper presented at the Annual meeting of the evaluation network, Austin, TX.

Stufflebeam, D.L., (1982). A next step: Discussion to consider unifying the Evaluation and research society (ERS) and Joint committee standards. *New Directions for Program Evaluation: Standards for evaluation practice*. 15, 27-36

Suchman, E. A. (1967). *Evaluation Research: Principles and practice in public services and social action programs*. New York: Russel Sage Foundation.

Taut, S. (2000). *Cross-Cultural Transferability of the Programme Evaluation Standards*. In The Programme Evaluation Standards in International Settings. Ed. By Russon C. The Center for Occasional Papers Series.

The Easy African Standard. (2002, April 13<sup>th</sup>). On 8-4-4 Education System of Education.

Thomas, R.M. (1992). Education's role in national development: Ten country cases. New York: Praeger Publishers.

Thompson, A.R. (1985). Evaluation in overseas education: Some parameters and priorities. *International Journal for Education Development*. 5(3), 227-233.

Turnbull, B. (1999). The mediating effect participation efficiency on evaluation use. *Evaluation and Programme Planning*. 22, 131-140.

Tyler, R. (1942). General statement on evaluation. *Journal of Education Research*. 35, 492-501.

UNESCO, (1994). Evaluation of program activities at UNESCO: Central program evaluation unit. Paris.

Watson, K. (1988). Forty years of education and development: From optimism to uncertainty. *Education Review*. 40, 137-174.

- Weiss, C.H., (1987). Evaluating social programmes: What have we learned? *Society*. 25, 40-45.
- Wholey J. S. (1987). Using evaluation to improve program performance. *Evaluation Studie*, 12, 55-69.
- Wilcox, T.G. (1989). Programme evaluation in education. Published thesis: Alberta University of British Columbia, British Columbia, Canada.
- World Bank. (1991). World development report: The challenge of development. New York: Oxford University Press.
- World Bank. (2000). Monitoring and evaluating capacity development in Africa. Selected proceedings from a seminar and workshop organised by the Development Bank of South Africa, the African Development Bank and the World Bank. Johannesburg: South Africa
- Worthen, B. R. (1994). Is Evaluation a Mature Profession that Warrants the Preparation of Evaluation Professions. *New Directions for Programme Evaluation*. 62, 3-15.
- Worthen, B.R., & Sanders, J.R., (1973). *Education evaluation. Theory and practice*. Belmont, CA: Wadsworth.

Worthen, B.R., & Sanders, J.R. (1987). Education Evaluation: Alternative approaches and practical guidelines. New York: Longman.

Worthen, B.R., Sanders, J.R. & Fitzpatrick, J.L. (1997). Programme Evaluation: Alternative Approaches and Practical Guidelines. New York: Longman.

## **APPENDIX : A**

## Appendix A :

### The Programme Evaluation Standards

Established by 16-professional education associations, identify evaluation principles that when addressed should result in improved programme evaluations.

Chaired by *Dr. James R. Sanders*; The Evaluation Center; Western Michigan University; Kalamazoo, Michigan 49008-5178 ; 616-387-5895.

Approved by the *American National Standards Institute* as an American National Standard. Approved date: March 15<sup>th</sup>, 1994.

*This is not copyrighted material. Reproduction and dissemination is encouraged.*

#### Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

- U1 **Stakeholder Identification** \_ Persons involved in or affected by the evaluation should be identified so that their needs can be addressed.
- U2 **Evaluator Credibility** \_ The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.
- U3 **Information Scope and Selection** \_ Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.
- U4 **Values Identification** \_ The perspectives, procedures and rational used to interpret the findings should be carefully described so that the bases for value judgement is clear.
- U5 **Report Clarity** \_ Evaluation reports should clearly describe the program being evaluated, including it's context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.
- U6 **Report Timeliness and Dissemination** \_ Significant interim findings and evaluation reports should be disseminated to intended users, so that, they can be used in a timely fashion.
- U7 **Evaluation Report** \_ 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.

#### Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic and prudent.

- F1 **Practical Procedures** \_ Evaluation should be practical, to keep disruption to a minimum while needed information is obtained.
- F2 **Political Viability** \_ The evaluation should be planned and conducted in anticipation of different positions of various interest groups, so that their cooperation maybe obtained, and that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.
- F3 **Cost Effectiveness** \_ The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

#### Propriety Standards

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 it's results.

- P1 **Service Orientation** \_ Evaluation should be designed to assist organisations to address and effectively serve the needs of the full range of targeted participants.
- P2 **Formal Agreements** \_ Obligations of the formal parties of an evaluation of (what is to be done, how, by whom, when) 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.
- P3 **Rights of Human Subjects** \_ Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.
- P4 **Human Interactions** \_ Evaluations should respect human dignity and worth in their interactions and other persons association with an evaluation, so that participants are not threatened and harmed.
- P5 **Complete and Fair Assessment** \_ The evaluation should be complete and fair in its examinations of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.
- P6 **Disclosure of Findings** \_ 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 evaluations, and any other with expressed legal rights to receive the results.
- P7 **Conflict of Interest** \_ Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation process and results.
- P8 **Fiscal Responsibility** \_ The evaluation 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.

### Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth of merit of the program being evaluated.

- A1 **Programme Documentation** \_ The programme being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.
- A2 **Content Analysis** \_ The context in which the program exists should be examined in enough detail so that it's likely influence on the program can be identified.
- A3 **Described Purposes and Procedures** \_ The purposes and procedures of the information should be monitored and described in enough detail, so that they can be identified and assessed.
- A4 **Defensible Information Sources** \_ The sources of information used in program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
- A5 **Valid Information** \_ The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.
- A6 **Reliable Information** \_ The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.
- A7 **Systematic Information** \_ The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.
- A8 **Analysis of Quantitative Information** \_ Quantitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered.
- A9 **Analysis of Qualitative Information** \_ Qualitative information in an evaluation should be appropriately and systematically analysed so that evaluation questions are effectively answered.
- A10 **Justified Conclusions** \_ The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.
- A11 **Impartial Reporting** \_ Reporting Procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that the evaluation reports fairly reflect the evaluation findings.
- A12 **Metaevaluation** \_ The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that it's conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

Guidelines and illustrative cases to assist evaluation participants in meeting each of these standards are provided in the *Program Evaluation Standards* (Sage, 1994).

## **APPENDIX : B**



## APPENDIX : B

### Format for Applying the Standards

#### ➤ UTILITY STANDARD (U)

Are intended to ensure that an evaluation will serve the information needs of intended users

#### U1 Stakeholder Identification

- Are persons involved in or affected by the evaluation identified
- Are both important and less powerful groups included
- Is there an attempt to reach an understanding with the client
- Are the clients involved in the designing and conducting the evaluation
- Is there any attempt to reach an understanding with the client
- Is there any discrimination in choice regarding the gender, ethnicity, language
- Are the objectives of the evaluation consistent with the needs of the audience
- Any inappropriate restriction on the evaluator in contacting the clients
- Can all the stakeholder information be needs be addressed

#### U2 Evaluator Credibility

- Is the person(s) conducting the evaluation both trustworthy and competent
- Do they exhibit the training, technical competence, substantive knowledge, experience
- If a team is involved do they collectively possess these qualifications
- Do the evaluator(s) maintain a pattern of consistent, open, and continuing communication
- Can they defend utility, practicality, integrity and technical adequacy
- Do they stay abreast of social and political forces with evaluation
- Is the work plan and composition of the team to concerns of stakeholders
- Is the audience informed about the progress of the evaluation
- Does the evaluation proposal include evaluator(s) qualification
- Does the evaluator determine if their approach is acceptable to the client

#### U3 Information Scope and Development

- Is the information selected pertinent questions of the programme

- Is the information responsive to the needs of the clients and stakeholders
- Is the evaluation relevant to the decision makers objectives
- Is the information important to significant stakeholders
- Is the information significantly comprehensive to support judgement of worth
- Is the information required addressing issues including that of effectiveness, harmful side effects, costs, response to learner needs, meaningfulness of assumptions, values underlying the programme, whether or not the stakeholders ask for this information.
- Are the questions to be addressed prioritized
- Has the evaluator reviewed pertinent literature, previous evaluations, research reports and discussions with stakeholders to provide background knowledge.
- Are additional info and anticipated info that emerge included

#### **U4 Values Identification**

- Are the perspectives, procedures and rationale used to interpret the findings described
- Is the interpretation done within defensibility of what has and has no merit
- Who is responsible for making the value judgement and determining the procedure for
- Are any alternative bases for interoperation provided e.g programme objectives, laws and regulations, institutional goals, democratic ideas, social norms, performance by comparison groups, needs of consumer, professional standards, judgement by various reference groups
- Any use of alternative techniques in designing value meanings e.g advocacy reports
- Allowing debate in the process as decision rules are arbitrary

#### **U5 Report Clarity**

- Is the evaluation report clearly described to include context purpose, procedures, findings
- Do the stakeholders readily understand the entire evaluation process from initial stage of purpose, the evaluand, process used, information obtained, conclusions drawn, recommendations, and descriptions.
- What forms do the reports take \_ oral, written, documents, video, graphic
- Are there special reports for special and different audience \_ clients and stakeholders
- Is there sufficient contextual info to back conclusions and recommendations
- Is the technical language such as validity, reliability purposive sampling backed by glossary, separate summary
- Are summary statements supported with discussions of related problem\*
- Are reports presented for review by client and intended audience
- Are evaluation reports translated through oral explanations or translations at meetings
- Is methodology overemphasized at the expense of findings
- Is the programme described adequately in the report

## U6 Report Timeliness and Dissemination

- Are interim findings and evaluation reports disseminated to intended users \_ client who commissioned the evaluation, legally responsible persons, sponsors, substantial informants, those quoted in the reports such as parents, students, mass media,
- Are the reports used timely \_ at times when it can be best used
- Is special effort made to identify , reach and inform intended users
- Are the reporting approaches appropriate \_ executive summaries, printed reports, presentations : audio visual, meetings, conferences, interviews, panel discussions, news paper
- Is the authority to fulfill this requirement of dissemination specified in the formal agreement
- Are there any violations of such an agreement
- Does the client specify the reporting form and times for dissemination
- Is there sufficient time allowed to submit the results
- Is there independent audit of the report by persons other than the client or the evaluator
- To what extent are the evaluators sensitive to social impediments to disseminate, e.g religious, culturalness, social behaviours, language barriers

## U7 Evaluation Impact

- Is the evaluation reported in ways that encourage use to influence decisions
- Are the evaluators involved in helping in assessing, making constructive use of the results, and to understand the programme in new ways \_ P.S They should not replace the client
- Is there indication that stakeholders are told how the findings might be used at the beginning
- Is the evaluator open, frank, and concrete in reporting
- Is the merits and demerits of findings discussed with the stakeholders
- Is there a balance between theoretical value of the findings and practical value
- Is there failure to respond when results are misused, misinterpreted and withheld

### ➤ FEASIBILITY STANDARDS (F)

Are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

## F1 Practical Procedures

- How are the contractual agreement with the client reached
- How are data sources chosen

- Which instruments are used and how are they administered
- How is data and information collected, recorded, stored, retrieved, analysed
- How are findings reported
- To what extent are disruptions minimised given the time, budget, programme stoppage
- Preliminary preparations undertaken e.g training in process
- Is evaluation part of routine events or timed
- Are the try outs such as piloting for practicality, reliability and time requirements
- Is method chosen appropriate

## **F2 Political Viability**

- Is the evaluation planned and conducted with anticipation of different positions of various interest groups that seek to influence policy.
- Are the political implications such that they can be handled with fairness and equity
- Is there sensitivity to various controversial groups, and are they indicated
- Are groups threatened by the evaluations taken care of
- Have evaluators met with such groups to express their positions
- Does the contractual agreement make explicit conditions that will govern the evaluation
- Is there provision made for periodic revelation of reports to avoid total surprises
- Have there been any evaluations discontinued, and on what account
- To what extent is the right of the public and all else to know upheld

## **F3 Cost Effectiveness**

- Is the evaluation efficient and does it produce information of sufficient value
- Does the cost effectiveness benefits equal or exceed monetary, non-monetary, social, human, physical costs . Benefits are value of all results.
- Are there cases where one is mandated to carry out evaluations whatever the costs.  
How often
- Is there thorough investigation on initial costs of services and materials
- Is there a detailed budget developed with all costs, with time and in-kind support
- Is there an inventory of benefits agreed upon by clients
- Are evaluators generally prudent and efficient in expending resources

### **➤ PROPRIETY STANDARDS (P)**

**Are intended to ensure that an evaluation will conduct legally, ethically and due regard for the welfare of those involved in the evaluation, as well as those affected by it's results**

## **P1 Service Orientation**

- Does the evaluation serve participants, community and society
- Are education goals appropriate
- Is learner and teacher performance, development addressed \_ intended and not
- Are promised services delivered
- Is effectiveness of programme monitored
- Has there been recommendation to terminate programme \_ Reasons
- Are evaluation practices current with the state of the art

## **P2 Formal Agreements**

- Is obligations of what is to be done, how, by whom and when agreed upon in writing
- Where evaluation is internal is there a brief memorandum or meetings of agreement
- Indicate what would constitute a breach of contract and actions to be taken
- Is there collaboration with education administrators in drafting policies
- Are there cases when evaluator has acted unilaterally in matters where joint collaboration was agreed upon
- Any instances when parts of the agreement have been changed without amendment
- Are the contracts so detailed that they interfere with creativity of evaluation

## **P3 Rights of Human Subjects**

- Are evaluations designed and conducted to respect and protect human subjects
- Is there notification of rights and consent for participation
- Is there provision for privilege of withdrawal without prejudice and penalty
- Is there privacy of certain opinions, information, confidentiality, identity
- Is there considerations for ethical, common sense and courtesy \_ the right of ones physical and emotional preparedness, limits of time limits of involvement
- Does the evaluator understand the cultural and social values of the participants
- Are there instances when participants assigned control groups and denied treatment
- Is there appropriate written permission from authorities, subjects, parents, authorities
- Are the evaluators knowledgeable about these rights and privileges

## **P4 Human Interactions**

- Evaluators should respect human dignity and worth pertaining to interpersonal transaction
- Do the evaluators guard against potentially threatening or harmful effects
- Does the evaluator understand the culture, social values and language
- Are all the concerns of evaluation held by participants taken care of
- Do participants have an open communication channel with the evaluators
- Is any groups or persons given privileges where others are not
- Are findings on participants biased in any way due to personal reasons

- Does the evaluator discuss with others personal findings on participants
- Does the evaluator collect information that may embarrass the participant
- Does the evaluator on the other hand avoid embarrassing the participants by not addressing certain negative elements of the programme e.g unethical behaviour, fraud, waste, abuse etc

#### **P5 Complete and Fair Assessment**

- Is the evaluation complete and fair in its examination and recordings of strengths and weaknesses
- Are the positive and negative aspects of the programme fairly assessed and reported
- Are the processes used in the evaluation fully discussed
- Are intended and non-intended aspects reported
- Are there critical comments about strengths, weaknesses, thoroughness from knowledgeable
- re omissions due to time or cost constraints reported and estimated on overall judgment
- Is the reporting manipulated to please the partisan individuals or groups
- Does the reporting further or protect evaluators personal interests
- Is the reporting speculative or tentative
- Are limitations in conducting evaluations addressed fully
- Is the reporting one sided, that is, either negative or positive only

#### **P6 Disclosure of Findings**

- Are full set of evaluation findings along with pertinent limitations available to participants and others with legal rights to receive the results (except where the disclosure would endanger public safety or abridge individual freedom)
- Does the evaluation present frankly the evaluators judgement and recommendations
- Are those disclosing the results, public pronouncements and written reports open, direct and complete in their disclosure
- Do the persons and groups concerned have uncontrolled access to the results
- Are the results useful so that strengths can be built upon and problem areas addressed

#### **P7 Conflict of Interests**

- What conflicts occur when carrying out evaluations
- Are conflicts dealt with openly and honestly
- Are possible sources of conflict addressed
- Are procedures to protect against conflicts given in the agreement
- Are all stakeholders involved in the evaluation at all levels
- Are evaluation procedures, data, reports released publicly
- Are the advantages, gains and losses identified for all concerned
- Are there possibilities for metaevaluation
- Are persons uniquely qualified to be involved in the evaluation excluded/ included

## **P8 Fiscal Responsibility**

- Does the evaluators allocation and expenditure reflect sound accountability
- Are funds used for the purposes and procedures stated in the evaluation
- Is there any accounting and auditing procedures
- Is there indication for any misuse of funds
- Are major costs for the evaluation specified in the proposal and agreed upon
- Are there records kept of sources of funding and expenditure in a clear format

### **➤ ACCURACY STANDARDS (A)**

**Are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.**

#### **A1 Program Documentation**

- Is the programme described and documented clearly and accuracy
- Is reference made to characteristics such as personnel, cost, procedures, location, facilities, setting, activities, objectivity's, nature of participation, and potential side effects
- Are there separate descriptions for each aspect of the programme being studied
- Are there any discrepancies between intended and implemented

#### **A2 Context Analysis**

- Is the context in which the programme exists examined in enough detail that is the location, timing, political and social climate, competing activities, staff, pertinent economic conditions
- Sources from which the information is sort e.g, logs, records, demographic studies, news paper clippings, legislative bills, unusual circumstances, teachers strike
- Are the contextual factors described and reported or is the programme embedded and intertwined with it's context in ways difficult to disentangle
- Are any difficulties in doing the above included e.g divisive relationship in the administration, impoverished economic conditions

#### **A3 Described Purposes and Procedures**

- Are purposes of the evaluation stated in terms of evaluation objectives and intended uses of it's results
- Does the evaluation procedures include the ways in which the data and information are gathered, analysed, and reported to meet or satisfy the purpose

- Are points of agreement and disagreement identified and assessed
- Are the purposes described and reviewed at regular intervals of the evaluation
- Is there an account of the what was actually done
- Are purposes and procedures evaluated especially in large scale evaluations
- Are the purposes and procedures changed during the course of the evaluation
- See U5 \_ Report Clarity

#### **A4 Defensible Information Source**

- Is information obtained from a variety of sources for example, tests, surveys, observations (e.g classroom), interviews, documents, activities prior to testing
- Are these described (not just labeled), reported and criteria and methods used to decide on their choice given as part of the technical documentation
- Are the samples selected and the procedures used in sampling described and given
- Are any unique and biasing features of the obtained information
- Is there use of previously collected information
- Is hard quantifiable data given preference over interpretive data

#### **A5 Valid Information (Soundness and Trustworthiness \_ Validity)**

- Does the the evidence compiled support the interpretations and uses of the data and information collected from the given instruments and procedures
- Are the constructs (intelligence) and behaviours (degree of attainment) fully described
- Is there a description of the procedures implementation, judgement and scoring of responses/ observations, and how interpretations were made (U4 \_ Values Interpretation)
- Is there presentations of evidence \_ both qualitative and quantitative \_ that justifies their use
- Is there an overall assessment of all this in reference to the evaluation questions and answers to determine the degree of fit or congruence e.g F3 \_ Cost Effectiveness
- Are those involved in the evaluation credible and willing to perform tasks , U2 \_ Evaluator Credibility
- Is there any validity claims. Were they exploratory. Is the context taken into account \_ A2, Context Analysis). What about the qualifictaion of those involved.
- Is the characteristics of the respondents appropriate and considered e.g reading ability, language proficiency, disabled

#### **A6 Reliable Information (Degree of consistency of information \_ Reliability)**

- Are the sources of error present in each procedure considered and are steps to reduce or to describe the amount and impact of these unwanted sources on the results and findings
- Is the information collected directly relevant to the groups, and is the procedures appropiate
- Is rationale provided in relation to procedures, administering the instruments, and the heterogeneity of the groups



- Is the evaluator's expectation recorded as a check. Is there sensitivity towards perspectives of the stakeholders.
- Is there consistency procedures applied e.g inter -rater reliability \_ A11, Impartial reporting
- Is there training provided for data collectors as regards the instrument application
- Is the reliability of the Tests used established

#### **A7 Systematic Information**

- Is the information collected free from error as is possible, that is error from collecting, scoring, recording, coding, filing, collating, analysing, and reporting information
- Was the procedures for administering instruments followed and personnel trained \_ P1 \_ Service Orientation
- Is there a quality control check plan
- Is data entry verified by another person for accuracy

#### **A8 Analysis of Quantitative Information**

- Includes age, socio-economic characteristics, measures of achievement, attitude, behaviour, description, materials being evaluated
- Are the methods appropriate
- Is the analysis systematic that is, starting with exploratory analysis, followed by sophisticated and complex analysis plus visual displays
- Are different methods of analysis \_ multiple analysis\_ used to determine whether a replicable pattern
- Are differential effects for different groups taken into consideration
- Are the methods and approaches used defensible
- Are potential weaknesses in the research design, data analysis,, their influence on interpretations and conclusions reported

#### **A9 Analysis of Qualitative Information**

- Includes structured and unstructured interviews, participant and non-participant observations, hearings, documents and records
- Is qualitative information sort and appropriate procedures for collecting such information followed
- Is the analytic procedure and method for summarizing the data appropriate
- Are potential weaknesses reported e.g data from a single respondent
- Is the consistency of the findings safe guarded \_ inter -rater reliability

#### **A10 Justified Conclusions**

- Is the conclusions reached in the evaluation explicitly justified, defensible,

- Are conclusions based on pertinent information collected , sound analysis and logic, and accompanied by full information about how the evaluation was conducted
- Are there any plausible alternative explanations provided and explanations as to their rejection
- Do the conclusions respond to the stakeholders questions and questions of the evaluation and are they limited to these
- Is there any knowledge as to whether feedback was solicited from the participants
- Are audiences guided and advised to be cautious in interpreting equivocal findings
- Are possible side effects ignored in concluding
- Is conclusion based on insufficient or unsound information
- Is there too much caution displayed in interpreting the findings of the evaluation
- Are the limitation of the evaluation reported

### **A11 Impartial Reporting**

- Does the report reflect all the perspectives that should be taken into account
- Is the reporting influenced by personal feelings and biases of ant party
- Is there signs of pressure on the evaluator to distort reports
- Is there agreement reached with the client during the initial stages about the steps to be taken to ensure fairness of all reports
- Are alternative and conflicting conclusions and recommendations reported \_ A10, Justified Conclusions
- Is there an attempt to maintain independence of reporting by using techniques such as rotation of evaluation team members
- Are there steps taken to protect the integrity of the reports, by safeguarding against deliberate or inadvertent distortion
- Is there an attempt to please the client to the extent that the negative findings are not presented

### **A12 META-EVALUATION (Evaluation of Evaluation)**

- Is the programme evaluation itself evaluated, that is, is the process and procedures applied worthwhile, and meritocracy of a completed evaluation.

## **APPENDIX : C**

## Appendix : C

### Interview Schedule : Evaluation Context

#### **General Information:**

*Background information on the interviewee.*

Place of work:

Job Description:

Length of time in the field:

*Presentation of the subject:*

I am carrying out a study on programme evaluation of education programmes. The purpose is of the study is to gather information on the process of evaluation practice in primary and secondary in Kenya.

I would like to ask some questions about curriculum evaluation.

However, you do not have to answer questions you do not feel comfortable with.

If you would like to go back to any of the questions, please feel free to do so.

The interview will be taped for later reference and for accurate description of the outcome. Should you have any questions to ask about this, you may do so. The tapes will be erased after completion of the study.

Information collected will be treated with strict confidentiality.

If you have any questions you wish to ask, you may do so.

## Interview questions.

### Section : A : Establishing evaluation context in Kenya.

Officials required to answer the questions in this section are those from:

Ministry of Education personnel (permanent secretary, director of education etc), head teachers, school teachers, KIE personnel, university scholars, donor agencies and education consultants.

#### *Questions to ask in order to establish the mandate:*

- 1 What situations or contexts make evaluations necessary (request from the governing bodies; donor agencies; economic situation, special problems)?
- 2 Which individuals or groups interested in the evaluation results must be taken into account (beneficiaries (community, schools donors, member states with stakes in the program)?
- 3 What questions (priority ones) does the evaluation ask in order to :
  - 3.1 meet the needs of users and beneficiaries?
  - 3.2 help the principal client make decisions and meet his or her information needs?
  - 3.3 take specific interests into account e.g those of other stakeholders?
- 4 Who conducts the evaluation (professional evaluators, those who run programmes at local levels...)?
- 5 Who should be asked to carry out the evaluation and in what capacity, for example collection of data, analysis..)
- 6 What particular conditions and procedures are observed in conducting the evaluation (constraints linked to political, administrative, duties of parties involved...)?
- 7 How are programmes implemented? Is the information documented and are documents accessible? Which interested parties are involved?

*Questions to ask in order to prepare the evaluation:*

- 8 From what source is information about programmes obtained (reports, evaluations already conducted, programme specialists, beneficiaries...)?
- 9 What information is already available that could be useful in answering questions in part or whole and to what extent is the information at hand valid or reliable.
- 10 How are the programmes organised (distributions of roles and responsibilities, allocation of budgetary resources...)?
- 11 What kinds of information collection methods would make it possible to obtain the necessary quality in the information requested?
- 12 Can information be collected in accordance with the terms of reference drafted for the implementation of the evaluation exercise.
- 13 What are the specifications for preparing the report and its annexes?
- 14 What particular conditions are to be observed to ensure optimum utilisation of the results obtained \_\_
  - Special collaboration among various interest groups.
  - Special recommendations to determine what procedures to use for circulating the results.
  - Who is responsible for circulating the results and what procedures do they use.
- 15 Is there any assessment of the evaluation carried out?
- 16 What are the consequences of conducting evaluations?
- 17 To what extent are evaluation results used as expected?

***Evaluation procedures followed when conducting evaluating:***

18 Summative evaluation:

20.1 At what stage of the programmes is the evaluation conducted.

20.2 On what grounds is it decided.

19 Formative evaluation:

21.1 At what stage of the programmes is the evaluation conducted.

21.2 On what grounds is it decided.

**Section : B Policies related to programme evaluation in Kenya**

Officers to answer these questions are those in the government departments and ministries and donor co-operates at policy formulating level.

1 What are the organisation's evaluation policies in general.

1.1 The past, present and future policies.

1.2 Reasons for changes in the approach if any.

1.3 Course of action.

2 What are the policies specific to education.

2.1 The past, present and future.

2.2 Reasons for changes in the approach if any.

2.3 Course of action.

***Questions to be answered by Kenyan officers.***

3 Are you aware of the donor's mandate in respect to their policies as it relates to:

3.1 Evaluation specifically and evaluation improvement in particular?  
How would you summarise it.

3.2 To what extent is donor mandate compatible with the countries evaluation objectives.

4 State in your own words (three) of the objectives of evaluation.

5 Name some organisations that the government for evaluation.

collaborates with enumerating some of the programmes.

- 6 Can you name some programmes that have been discontinued by any donor agency. What were the reasons for discontinuing programme/project/collaboration

**The End..**



## **APPENDIX : D**

## Appendix : D

### Written Consent Form

I understand that you are carrying out a study on programme evaluation practices in education.

I volunteer to participate in the study on condition that information give and comments make will be treated confidentially.

*Please pick one:*

consent to my comments being quoted in the research report provided my identity protected and my name is not used.

do not consent to my comments being quoted in the research report.

Being fully aware of the nature of this study, I hereby agree to participate in it. I have received a copy of this consent.

\_\_\_\_\_  
Signature \_\_ Respondent

\_\_\_\_\_  
Date:

\_\_\_\_\_  
Respondent's name (print)

\_\_\_\_\_  
Signature \_ Researcher

## **APPENDIX : E**

## Appendix : E

### Contact sheet

#### *Particulars of person contact:*

Name: \_\_\_\_\_ Telephone No:

Address:

Date:

Place of work:

Job title and description:

#### *Introduction of interviewer:*

My name is

I am currently working at

I am doing a research on "Evaluation practices in education in Kenya." The purpose of the research is to gather information on the process of evaluation practices in primary and secondary school programmes in Kenya.

Could you please provide me with a list of programmes that have been evaluated previously and those currently undergoing evaluation. Also needed is the people who were key players in the programmes stated. The people stated should be familiar with the programme and/ or are directly involved.

would you be interested in taking part?

Yes \_\_\_\_\_ (Move to the next section)

No \_\_\_\_\_ (end of conversation and thank you)

will send you written confirmation of the arrangements.

If you wish to have more information, you may call me at:

Thank you for your Co-operation.

## **APPENDIX : F**

## Appendix : F

### Research Data Base

#### Documents reviewed

Sixteen documents were used for studies. These were labeled 001 to 016.

#### Data Source 1 : Document Composition

F: Formative evaluation, internally conducted (11) : 68.8%  
S: Summative evaluation, externally conducted (2) : 12.5%  
N: Needs assessment, internally conducted (3) : 18.7%

Identification No:	F	S	N
001	✓		
002	✓		
003	✓		
004	✓		
005	✓		
006	✓		
007	✓		
008	✓		
009	✓		
010	✓		
011	✓		
012		✓	
013		✓	
014			✓
015			✓
016			✓
Formative: 68.8% Summative: 12.5% Needs Ass: 18.7%			
Total : (16) : 100			

## Data Source No: 2a : Statements that define Purpose for evaluation

This information was derived from the objectives set out for the evaluations.

Identified purposes for the evaluation	%
<p><b>Formative evaluation:</b></p> <p><b>I. Aims and Objectives:</b></p> <ul style="list-style-type: none"> <li>• <b>Development and Improvement</b></li> </ul> <p><b>Formative:</b></p> <ul style="list-style-type: none"> <li>- Recommend procedures for strengthening education in Kenya (F)</li> <li>- How implementation problems could be remedied (F)</li> <li>- Identify problems in education and to improve them (NF)</li> <li>- Collect info that will facilitate the review of the curriculum (F)</li> <li>- Collect data for use in planning in-service courses (F)</li> <li>- Set directions for further development of school education programme (F)</li> </ul> <p><b>Summative:</b></p> <ul style="list-style-type: none"> <li>- Make recommendations for school improvement (NS)</li> </ul> <p style="text-align: right;"><b>Total :</b></p>	<p>(6) : 19.4</p> <p>(1) : 3.2</p> <p>(7) : 22.5</p>
<ul style="list-style-type: none"> <li>▪ <b>Awareness and knowledge</b></li> </ul> <p><b>Formative:</b></p> <ul style="list-style-type: none"> <li>-Extent to which school curriculum achieves objectives of school (NF)</li> <li>-Extent to which school curriculum has been effectively implemented(NF)</li> <li>-Identify main strengths and weaknesses of education to improve it(F)</li> <li>-Determine effectiveness of in-service courses (NF)</li> <li>-Identify problems teachers are facing in implementation of curriculum (F)</li> <li>-Determine reactions of pupils, teachers and parents towards the curriculum (NF)</li> <li>-Identify and resolve scope and sequence issues of subject areas (F)</li> <li>-Document programme to facilitate on-going revision (F)</li> <li>-Provide feedback to curriculum developers (F)</li> <li>-Determine teachers competency (F)</li> </ul> <p><b>Summative:</b></p> <ul style="list-style-type: none"> <li>-Determine extent to which objectives of education are realistic and achievable (S)</li> <li>-To assess the relevance of the school curriculum (S)</li> <li>-To find out the efficiency, effectiveness, and effects of the curriculum development (S)</li> <li>-To assess the efficiency of the implementation process (NS)</li> <li>-To find out the attitudes of the learners, teachers and parents towards curriculum (S)</li> </ul>	<p>(10) : 32.2</p> <p>(5) : 16.2</p>



Data source No: 2a cont.....//

**Needs Assessment:**

Provide information on:

- National goals of education
- Objectives of schools
- Appropriate curriculum content of schools
- The cost effectiveness of the curriculum with regard to available resources
- Developing programmes on social, moral, and technological development
- How secondary school curriculum can enhance industrial development

(7) : 22.6

▪ **Decision Making**

- Provide timely information for decision making

(1) : 3.2

▪ **Planning**

- Provide information to assist in planning implementation

(1) : 3.2

(F) : Formatively stated

(NF) : Not Formatively stated

(S) : Summatively stated.....(NS) : Not Summatively stated

**Data Source No: 2b : Summary occurrence of commonly found purposes:**

Purpose	F	S	N	Total %
Development and Improvement	20%	3%	0%	22.6%
Awareness and knowledge	32%	16%	23%	71.0%
Decision Making				3.2%
Planning				3.2%
<b>Total (31) : 100%</b>				

**Data Source No: 2c : Orientation of the stated purposes to determine direction of focus**

Nature of Evaluation	Yes	No	Total%
Formatively stated	(11) : 68.8	(5) : 31.2	(16) : 100
Summatively stated	(4) : 66.7	(2) : 33.3%	(6) : 100

### Data Source No: 3 : Reasons for undertaking the evaluation

(The decision was made on who desired the evaluation.)

Identification No:	Reason?
001	A
002	A
003	A
004	A
005	A
006	A
007	A
008	A
009	A
010	A
011	A
012	P
013	P
014	P
015	P
016	P
Administration : (11) : 68.7% Policy : (5) : 31.3% Total : (16) : 100%	

P : Government \_ Policy making body

A : Curriculum advisory body \_ K.I.E

## Data Source No: 4a : Commonly occurring objects of evaluation

(Derived from stated purposes and objectives : 10-documents with highest congruence were used)

Categories and Objects of Evaluation	Formative	Summative	Needs Assessment	Total %
<ul style="list-style-type: none"> <li>▪ Education/School Philosophy</li> <li>▪ Curriculum Programme</li> </ul>	IIII IIII	IIII	III	(17) : 25.8
<ul style="list-style-type: none"> <li>Philosophy and Objectives</li> <li>Development</li> <li>Content</li> <li>Efficiency and costs</li> </ul>	IIII— III	II II	I I I I	(17) : 25.8
<ul style="list-style-type: none"> <li><b>Rel. to Industrial Development</b></li> <li>Inclusion _ Social and Moral science</li> </ul>			I I	I
<ul style="list-style-type: none"> <li>▪ Subject area</li> <li>Scope</li> <li><b>Sequence</b></li> </ul>	IIII— III—		I I	(8) : 12.1
<ul style="list-style-type: none"> <li>▪ Personnel</li> <li>▪ Learners needs</li> <li>▪ Implementation</li> <li>▪ Reaction :Teachers, Pupils, Parents Community members</li> </ul>	IIII IIII— III	I II II	II I I	(7) : 10.6 (9) : 13.6 (8) : 12.1
<p>Number of objectives:  <b>TOTAL : 66 : 100%</b></p> <p><b>Formative evaluations : (37) : 56.1%</b>  <b>Summative evaluations: (17) : 25.7%</b>  <b>Needs Assessment: (12) : 18.2%</b></p>				

The object of the evaluations were derived from the stated objectives.

**Data Source No: 4b : Summary occurrence of commonly found objects of evaluation:**

Objects	Total %
Education/ School Philosophy	25.8
Curriculum Programme	25.8
Implementation	12.1
Subject Area _ Course units taught	10.6
Personnel	13.6
Reactions	12.1
Inclusion of Social and Moral Science	12.1
<b>(66) : 100%</b>	
	<b>100%</b>

**Data Source No: 5a :**

**Information sources, Approaches and Techniques applied in the evaluations**

Approaches / Techniques/ Informants	F	S	N A
<b>Qualitative:</b>			
Interview schedules			
<b>School Level</b>			
-Students	-----	II	---
-Teachers	-----	II	---
-Parents	IIIIIIII--	II	III
-Headteachers	-----	II	II--
-BOG	IIIII--	II	---
Community leaders	IIIIIIII--	II	III
Religious leaders	-----	II	
<b>MOE:</b>			
-Education officers	-----	--	
-Inspectorate	III--	II	II--
-K.I.E	II--	II	---
-K.N.E.C	-----	II	---
-T.S.C	-----	--	---
-K.N.U.T	-----	II	---
Religious groups	-----	II	---
Special needs (the disabled)			---
<b>Out of School:</b>			
School leavers:		II	

Dropouts: -Other	----- -----	II	---
Focus Group Discussions	-----	II	III
Content Analysis	III-----	II	III
Document Reviews	III-----	II	II-
<b>Quantitative:</b>			
Achievement Tests (Student outcome)	II---	--	--II
Questionnaires	IIIIII	II	--II
-Students	IIIIII	-	--II
-Teachers	-----	-	---
-Parents	-----	II	I---
-Headteachers	---	--	--II
<b>Attitudinal scale</b>			
Observation schedule			
-Classroom process	III	II-	--II
-Lesson Plan	III	II-	--II
Check list (school records/ environment)	III	II II	III

**Data Source No: 5b : Summary of Instruments used to elicit Information**

Data Collection instruments	F	S	NA	Total%
Interviews	18	26	12	44.1
Questionnaires	12	4	4	15.7
Observation Schedules	8	4	2	11.0
Check lists	6	2	4	9.4
Content Analysis	2	2	4	6.4
Document Reviews	-	2	4	4.7
Attitudes	-	-	2	1.6
Student outcome measures	2	-	1	4.7
Source =The sixteen evaluation review documents				
				100%

## Data Source No: 6a : Units of Analysis \_ Formative evaluation

Derived from the designs and the analysed data presented

Units	Cross Cutting Issues
<b>Formative:</b>	
School types:	Day, boarding, mixed, single, national,
Region:	Rural/urban;
Class levels:	Grades, sizes, learning process
Course units:	Sequence, context, time allocation, relevance to objectives, process, Text Books  Towards school, 8-4-4
Attitudes:	Teachers, HeadTs,' Learners,
Respondents:	Community

.....Data Source No: 6b : Summative evaluations

Units of Analysis	Cross Cutting Issues
<b>Summative:</b>	
School types:	Day, boarding, mixed, national, Boys / Girls, Year of establishment
Region:	Urban/Rural; Province/ District;
Class levels:	Grades, sizes, learning process, gender
Course units:	Sequence, context, time allocation,
Attitudes:	Towards school, 8-4-4
Respondents: -Teachers	Province, gender, age, Academic and profession
Head Teacher	Province, Gender, sex, age, academic & profession

**Data Source 7a : Summary representation of modes of analysis**

Types	Categories	F	S	NA
Needs Assessment				
Document reviews	Education Commission reports, Past evaluations	✓	✓	✓
Quantitative: ✓	Descriptive: Percentages, Quantities	✓ ✓	✓	✓ ✓
Qualitative:	Analysis of Variance: T-Test			✓ ✓
	Categorised interpretations Number of occurrence Categories of descriptions			✓ ✓ ✓
Presentation:	Tables	✓	✓	✓

**Data Source No: 7b : Sample presentations of data analysis from evaluation documents**

▪ **Sample 1**

In response to the reason why students did not like the secondary school course:

Response	Percentages %
Too much work	30
Lack of Text Books	29
Poor Teaching	25
Lack of teachers	16



### Sample 2

Student Achievement : mean Scores and Standard Deviation

Subject	Form I	
	Mean (x)	SD
Kiswahili	60	13
Mathematics	18	17.2
Chemistry	16.2	9.2
History/ Geography	49.5	15
Social Education and Ethics	73	16
Home Science	24.6	9
Geography	31.2	13.2

### Sample 3

In response to the curriculum work load?

Response	Form I %	Form II%	Reason
YES	92.01	84.89	Too many subjects Syllabus loaded
NO	3.05	3.4	--

### Sample 4

In terms of in service carried out?

In-service	Kiswahili %	English %	Biology %	Physics %
Non	80	26	44.4	37
1	8.0	15	22.2	29
2	8.0	17	19.4	7
3	4.0	9	16.7	
4	0	5	0	
5	0	11	0	

## Data Source No: 8a : Nature and Quality of outcome of evaluations

T : Tentative \_ Not carefully thought out

S : Speculative \_ Carefully thought about

T/S : Neither Tentative or Speculative

+ve : Positive reporting

-ve : Negative Reporting

Category of Evaluation	Nature of Outcome	S/T	+ve -ve
Formative	<b>Education and School Philosophy:</b>	-	-
	<b>Curriculum Programme:</b> Broad and overloaded curriculum Syllabus could not be completed within time Content too wide	T S T	-ve -ve n
	<b>Subject Area:</b>		
	<b>Personnel:</b> Inadequate in-service for teachers Lack of teaching resources	S S	- ve -ve
	<b>Implementation:</b> Problems with implementation Inadequate resources There is lack of facilities and resources	T S S	-ve -ve -ve
	<b>Reactions:</b> Attitude towards secondary school is negative	S	-ve
Summative	<b>Education and Philosophy:</b> School objectives are relevant and desirable	S	n
	<b>Curriculum Programme:</b> Curriculum objectives are too broad Curriculum not easily achievable Syllabus very broad Content areas should be reduced Curriculum development needs to be made efficient	T T T S/T T	n -ve -ve n n
	<b>Subject Area:</b> Pastoral as a subject should be scrapped	S	n

	<p><b>Personnel:</b>  Teachers found syllabus difficult to complete  More time was needed for teacher orientation  One week training for subject areas was too short</p> <p><b>Implementation:</b>  Supply of textbooks was found to be inadequate  There was inadequate supply of equipment  Primary class sizes _ no: of students should be reduced to 30</p> <p><b>Reactions:</b>  Students attitudes were:  Positive towards traditional subjects and  Negative towards practical subjects  Teachers had a positive attitude towards the curriculum</p> <p>Curriculum prepares graduates for further education and employment</p> <p><b>Policy Level:</b>  K.I.E should have been involved in the in-service  Examinations were appropriate to the syllabus  Awareness creation was not well done</p>	<p>S/T S S/T</p> <p>T T S</p> <p>S/T S S/T</p> <p>S/T</p> <p>S S S/T</p>	<p>-ve n -ve</p> <p>-ve -ve n</p> <p>+ve -ve +ve</p> <p>n</p> <p>n +ve -ve</p>
	<p><b>Curriculum Programme:</b>  Curriculum is overloaded  Syllabus is too wide to be completed in time  Overlaps in curriculum  Too many subjects</p> <p><b>Resources and materials:</b>  There is inadequate resources to implement the curriculum</p> <p><b>Personnel:</b>  Teachers are inadequately trained</p> <p><b>Other:</b>  Parents cannot afford education</p>	<p>T S/T S S/T</p> <p>S</p> <p>S</p> <p>S/T</p>	<p>n n n n</p> <p>-ve</p> <p>-ve</p> <p>-ve</p>
	<p>Summary Analysis /// continued next page</p>		

<b>Summary of analysis</b>		
<b>S: Speculative</b>	<b>: (25)</b>	<b>: 55.6%</b>
<b>T: Tentative</b>	<b>: (20)</b>	<b>: 44.4%</b>
<b>Total</b>	<b>: (45)</b>	<b>: 100%</b>
<b>-ve : Negative Reporting</b>	<b>: (19)</b>	<b>: 54.3%</b>
<b>+ve : Positive Reporting</b>	<b>: (3)</b>	<b>: 8.6%</b>
<b>n : Neutral Reporting</b>	<b>: (13)</b>	<b>: 37.1%</b>
	<b>Total</b>	<b>: (35) : 100%</b>

**Data Source No: 8b : Summary analysis of quality of outcomes**

	<b>Speculative</b>	<b>Tentative</b>
<b>Formative</b>	<b>(6) : 13.0%</b>	<b>(3) : 6.5%</b>
<b>Summative</b>	<b>(20) : 43.5%</b>	<b>(17) : 37.0%</b>
<b>Total (46) : 100%</b>	<b>(26) : 56.5%</b>	<b>(20) : 43.5%</b>

**SP: Speculative:** Carefully thought about

**T: Tentative:** Not carefully thought about

**Data Source No: 8c Summary of negative and positive statements**

	<b>-ve %</b>	<b>+ve %</b>	<b>Neutral %</b>
<b>Formative</b>	<b>(8) : 22.9</b>	<b>(0) : 0.0</b>	<b>(1) : 2.8</b>
<b>Summative</b>	<b>(11) : 31.4</b>	<b>(3) : 8.6</b>	<b>(12) : 34.3</b>
<b>T : (35) : 100%</b>			
	<b>(19) : 54.3</b>	<b>(3) : 8.6</b>	<b>(14) : 37.1</b>

**Data Source No: 9a : Coverage of object of evaluation**

(Derived from statements of evaluation results/outcomes)

Category	Outcome	Coverage %
Summative	<p><b>Education and School Philosophy:</b></p> <p><b>Curriculum Programme:</b> Broad and overloaded curriculum Syllabus could not be completed within specified time Content too wide</p> <p><b>Personnel:</b> Inadequate in-service for teachers Lack of teaching resources</p> <p><b>Implementation:</b> Problems with implementation Inadequate resources There is lack of facilities and resources</p> <p><b>Reactions:</b> Attitude towards secondary school is negative</p> <p><b>Subject Area:</b></p>	<p>(3) : 8.3</p> <p>(2) : 5.6</p> <p>(3) : 8.3</p> <p>(1) : 2.8</p> <p>T : ( 9 ) : 25%</p>
Formative	<p><b>Education and Philosophy:</b> School objectives are relevant and desirable</p> <p><b>Curriculum Programme:</b> Curriculum objectives are broad Curriculum not easily achievable Syllabus very broad Content areas should be reduced Curriculum development needs to be made efficient Curriculum prepares graduates for further education and employment</p> <p><b>Subject Area:</b> Pastoral as a subject should be scrapped</p> <p><b>Personnel:</b> Teachers found syllabus difficult to complete More time was needed for teacher orientation One week training for subject areas was too short</p>	<p>(1) : 2.8</p> <p>(6) : 16.7</p> <p>(1) : 2.8</p> <p>(3) : 8.3</p>

	<p><b>Implementation:</b>  Supply of textbooks was found to be inadequate  There was inadequate supply of equipment  Primary class sizes _ no: of students should be reduced to 30</p> <p><b>Reactions:</b>  Students attitude was :  +ve. Towards traditional subjects and  -ve. towards practical subjects  Teachers had a positive attitude towards the curriculum</p> <p><b>Policy Level:</b>  K.I.E should have been involved in the in-service  Examinations were appropriate to the syllabus  Awareness creation was not well done</p>	<p>(3) : 8.3</p> <p>(3) : 8.3</p> <p>(3) : 8.3</p> <p>T : (20) : 55.6%</p>
Needs Assessment	<p><b>Curriculum Programme:</b>  Curriculum is overloaded  Syllabus is too wide to be completed within agreed time  Overlaps in curriculum  Too many subjects</p> <p><b>Resources and materials:</b>  There is inadequate resources to implement the curriculum</p> <p><b>Personnel:</b>  Teachers are inadequately trained</p> <p><b>Other:</b>  Parents cannot afford education</p>	<p>(4) : 11.1</p> <p>(1) : 2.8</p> <p>(1) : 2.8</p> <p>(1) : 2.8</p> <p>T : (7) : 19.4%</p>
	<p>Formative : (9) : 25%</p> <p>Summative : (20) : 55.6%</p> <p>Needs Assessment : (7) : 19.4%</p> <p>Total : (36) : 100</p>	

**Data Source No: 9b: : Summary of the extent to which outcomes relate to information sort**

This was determined by looking into the extent to which the statements of outcomes reflect the object of evaluation proportionately.

<b>Object of evaluation</b>	<b>F %</b>	<b>S %</b>	<b>Total%</b>
Education Philosophy	1	2	(3) : 10.7%
Curriculum Programme	2	6	(8) : 28.6%
Subject Area	1	1	(2) : 7.2%
Personnel _Teacher	2	3	(5) : 17.9%
Facilitator	3	3	(6) : 21.4%
Reactions	1	1	(2) : 7.1%
Policy level	1	1	(2) : 7.1%
Learners needs	-	-	-
	<b>11</b>	<b>17</b>	<b>(28) : 100</b>

**Data Source No: 10 : Sources of criteria for judgment**  
 (Derived from Results and Outcomes)

Broad Categories	Nature of Criteria for judgement	Total %
<b>Policy and Administration</b>	Curriculum as a Programme, School type, Classroom size, Subject area, Learner, Teacher, Education philosophy Relevant of Curriculum <b>Other:</b> Enhancing Industrial	<b>(9): 34.6</b>
<b>Process:</b>	Efficiency of implementation, Planning, In-service, Assess relevance, Enhance Industrial development	<b>(5) : 19.2</b>
<b>Outcome measures:</b>	<b>Descriptive :</b> Identified problems, Provision Reviews, Provision of Information, Provision guidance, Make recommendations, Feed Back information, Documentation, Teacher competency, Develop programmes  <b>Opinion:</b> Attitudes _ school and community Reactions  <b>Student Achievement</b> Test Scores	<b>(12) : 46.2</b>
		<b>(26) : 100</b>





Targeted area	Continue	Modify	Innovate	Terminate
<b>Summative Evaluation:</b>				
<b>Curriculum Programme:</b>				
Objectives are too broad and not easily achievable	-	G	-	-
Should take into consideration special needs	-	G	-	-
Need to review the curriculum to made it manageable	-	-	S	-
Curriculum should be reviewed to compare with international standards.	-	S	-	-
	-	G	-	-
	-	-	S	-
<b>Objectives should include physical education</b>	-	G	-	-
Some objectives need to be made clear (2,3, 5, & 6)	-	G	-	-
Should respond to the needs of learners	-	-	S	-
Need to focus on cognitive, affective and psychomotor	-	-	S	-
Curriculum load should be reduced II	-	-	S	-
Curriculum choice should be made flexible	-	S	-	-
Social, Morals and Health education shd. Be included	S	-	-	-
Subjects that enhance technology development shd. be taught.				
Inspection should be regularly carried out				
Primary education should retain:	S	-	-	-
English, Maths, GHC, Science, Agriculture, Religious education, Kiswahili, Physical education				
Subjects in primary to teach fine art, music, arts, only where there is facilities:	-	-	-	S
	-	S	-	-
	-	G	-	-
<b>Subject Area:</b>				
Pastoral subject should be scrapped	-	S	-	-
Strengthen assessment of practical subjects	-	S	-	-
Need to remove unnecessary overlaps	-	S	-	-
Kiswahili language and literature need to be separated	-	G	-	-
English language and literature need to be separated	-	G	-	-
English and mathematics should form core area				
Regulate subject area and time allocated				
Need to reduce the number of subjects	-	G	-	-
<b>Materials and equipment:</b>				
Enhance training packages for teachers	-	S	-	-
There should be adequate facilities, equipment and materials	-	G	-	-
<b>School achievement:</b>				
Should be enhanced by e.g reducing curriculum load				
<b>Personnel:</b>				
	-	G	-	-
<b>Specific / non-specific:</b>				
There is need to train curriculum developers Teachers should be trained in Testing	-	S	-	-
	-	S	-	-
There is a need to in-service teachers	-	S	-	-
Ensure there is enough personnel for all subject areas	-	S	-	-
Head teachers should be trained	-	S	-	-

Train more teachers in technical areas	--	S	--	--
There should be equitable distribution of teachers in schools	--	G	--	--
Deployment should be equitable done in all schools	--		--	--
Primary TTCs should raise minimum entry level, improve content area,	--		--	--
<b>Programme Implementation:</b>	--	G	--	--
Need to improve curriculum implementation strategies	--		--	--
<b>Policy and administration:</b>	--	S	--	--
There is need to decentralise in-service training	--	S	--	--
Should expand K.I.E to develop school materials prod.	--	S	--	--
Reduce class sizes to 30 students	--	S	--	--
Orientation of teachers should be continuous	--	S	--	--
Teachers workload should be reduced from 45-50 to 30 lessons	--		--	--
Regulate mock exams to reduce costs	--	S	--	--
Need to increase number of national schools	--	S	--	--
Enforce policy on the number of required subjects	--	G	--	--
There should be no extra tuition after class hours	--	--	--	G
Need to strengthen guidance and counseling	--	S	--	--
Government should cover most of education cost	--	S	--	--
Existing schools should be expanded and refurbished	--	S	--	--
Introduce well equipped technical schools	--	S	--	--
There should be equity in terms of gender, religion, social	--	--	--	S
All school levies in primary school should be removed	--		--	--
<b>Decision Taken</b>				
<b>Total: 61</b>				
<b>G : 23 : 37.7%</b>				
<b>S : 38 : 62.3%</b>				
	(5) : 8.2%	(45) : 3.8%	(7) : 11.4%	(4) : 6.6%

**Data Source No: 11b : Decisions taken based on the recommendations**

(Derived from Data source No: 11a)

- Continue:** Leave things as they are.
- Modify:** Make changes for improvement.
- Innovate:** Introduce something new.
- Terminate:** End activity.

Decision Taken	Continue %	Modify %	Innovate %	Terminate %
	G: (3) : 4.9 S: (2) : 3.3	G: (28): 45.9 S: (17) : 27.9	G: (1) : 1.6 S: (6) : 6.5	G: (2) : 3.3 S: (2) : 3.3
<b>Total : (61) : 100%</b> G: (23) : 37.7% S: (38) : 62.7%				
	<b>T : (5) : 8.2%</b>	<b>T:(45) : 73.8%</b>	<b>T : (7) 11.4%</b>	<b>T : (4) : 6.4%</b>

**Data Source No: 11c : Summary of the quality of recommendations made**

- G: Generalises activity to act on
- S: Specifies the exact activity to act on

Statements made	General %	Specific %
<b>Formative</b>	(8) : 12.9	5 : 8.1
<b>Summative</b>	(15) : 24.2	(34) : 54.8
<b>Total : 41 : 100%</b>	<b>25 : 61%</b>	<b>16: 39%</b>

**Data Source 12a : Users and Targets of Evaluation results**

Derived from purposes set for the evaluations and the recommendations as stated

Form of Evaluation	Intended (Derived from Purposes) %	Affected (From Recommendations) %
<b>Formative: (Internal)</b>	<p><b>Decision Makers</b></p> <p>Policy _ MOE (3) : 15                      Programmers _ K.I.E (8) : 45</p> <p><b>Recipients</b></p> <p>Learners (1) : 5                      Teachers (6) : 30                      Parents (1) : 5                      Society/Comm. (1) : 5</p> <p><b>Total : (20) : 100%</b></p>	<p><b>Decision Makers</b></p> <p>Policy _ MOE (7) : 35                      Programmers _ K.I.E (9) : 45</p> <p><b>Recipients</b></p> <p>Learners (-) : 0                      Teachers (3) : 15                      Parents (1) : 5                      Society/Comm. (-) : 0</p> <p><b>Total : (20) : 100%</b></p>
<b>Summative: (External)</b>	<p><b>Decision Makers</b></p> <p>Policy _ MOE (3) : 27.2                      Programmers _ K.I.E (4) : 36.4</p> <p><b>Recipients</b></p> <p>Learners (1) : 9.1                      Teachers (2) : 18.2                      Parents (1) : 9.1                      Society/Comm. (-) : 0</p> <p><b>Total: (11) : 100 %:</b></p>	<p><b>Decision Makers</b></p> <p>Policy _ MOE (30) : 35.7                      Programmers _ K.I.E (25) : 29.7</p> <p><b>Recipients</b></p> <p>Learners (1) : 1.2                      Teachers (9) : 10.7                      Parents (-) : 0                      Society/Comm. (-) : 0                      Schools (10) : 11.9                      Headteachers (1) : 1.2                      MOE_Inspectors (2) : 2.4                      MOE_KNEC (2) : 2.4                      MOE_TSC (2) : 2.4                      MOE_TTCs (1) : 1.2                      Special needs (1) : 1.2</p> <p><b>Total: (84) : 100%:</b></p>

Data Source No12b : Summary of users, targets and whether intended or not intended

Users and Targets of Evaluation	Intended		Affected	
	S	F	S	F
<b>Decision Makers</b>				
Policy _ MOE	27.2	15.0	35.7	35%
K.I.E	36.4	45.0	29.7	45%
<b>Recipients</b>				
Programmers _				
Learnners	9.1	5.0	1.2	1%
Teachers	18.2	30.0	10.7	15%
Parents	9.1	5.0	0	5%
Society/Comm.	0%	-	0	0
Schools	-	-	11.9	-
Head Teachers	-	-	1.2	-
MOE_Inspectors	-	-	2.4	-
MOE_KNEC	-	-	2.4	-
MOE_TSC	-	-	2.4	-
MOE_TTCs	-	-	1.2	-
Special needs (Handicap)	-	-	1.2	-

S: Summative

F : Formative

UNIVERSITY OF NAIROBI  
EAST AFRICANA COLLECTION