EXPERT EVALUATORS' NATURALISTIC DECISION MAKING DURING EVALUATIONS OF HIV/AIDS HEALTH EDUCATION PROGRAMS

by

ANITA FRANCES ZGAMBO

(Under the Direction of Ikseon Choi)

ABSTRACT

The goal of this research was to understand expert evaluators' naturalistic decision making during evaluations of HIV/AIDS health education programs. The focus was to understand the following aspects: (1) the critical incidents that evaluators encountered; (2) the factors that influenced evaluators' naturalistic decision making; and (3) how evaluators made decisions. A multiple case study research design, the Critical Decision Method, phenomenography, and narrative analysis were used to guide the methodological design and analysis of this study. Two one-hour interviews with seven participants (with ten to thirty years of evaluation experience) were conducted to develop narratives, decision making timelines, and models that captured evaluators' decision making processes. Klein's (1997) and Kundin's (2008, 2010) research on naturalistic decision making provided a framework for this study.

The major findings in this study are threefold. First, the evaluators encountered critical incidents at various stages in the evaluation life cycle, some of which acted as barriers and others that served as facilitators for performing evaluations. These barriers included challenges in the areas of communication, community entry, coordination, contextual understanding, and evaluation design. Evaluators encountered events that undermined their ability to implement an

evaluation, including data, budgetary, and time constraints, as well as resistance to evaluation. They also encountered events that facilitated their evaluations, such as successful stakeholder collaboration and sufficient time. Second, evaluators' naturalistic decision making was shaped by context, evaluation, human, and real-world decision making factors. Third, evaluators relied on situation assessments to understand the situations framing the evaluations. Their philosophical assumptions, interpersonal skills, self-determination, and use of satisficing assisted evaluators during the decision-making process.

Although expert evaluators attempted to anticipate barriers when implementing evaluations, they could not anticipate all the barriers they encountered. However, they adapted their evaluation strategies in response to these unexpected barriers. This reveals an opportunity to study evaluators as innovators, which may provide an insight into how evaluators innovate when they encounter unexpected barriers. Evaluators also established participatory relationships with stakeholders and engaged them in the evaluations. This is another opportunity to extend this research, thereby exploring how expert evaluators establish relationships with stakeholders.

INDEX WORDS: Experts, Evaluators, Naturalistic Decision Making, HIV/AIDS health education, Program evaluation, Critical incidents

EXPERT EVALUATORS' NATURALISTIC DECISION MAKING DURING EVALUATIONS OF HIV/AIDS HEALTH EDUCATION PROGRAMS

by

ANITA FRANCES ZGAMBO

BS, Kennesaw State University, 2004

MS, Kennesaw State University, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

© 2012

Anita Frances Zgambo

All Rights Reserved

EXPERT EVALUATORS' NATURALISTIC DECISION MAKING DURING EVALUATIONS OF HIV/AIDS HEALTH EDUCATION PROGRAMS

by

ANITA FRANCES ZGAMBO

Major Professor:	Ikseon Choi
Committee:	Michael Orey
	Thomas C. Reeves
	Lorilee Sandmann

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012

DEDICATION

In memory of my mother, Fabiola E. Romero, who instilled in me a love of learning and emphasized the importance of earning an education.

ACKNOWLEDGEMENTS

I would like to thank my family, my husband, Thomas, and our children, Frances-Feliz, Thomas, Patrick, and Wanangwa, for their encouragement and support. I am grateful for my adviser's support, especially for your valuable feedback and confidence when I became discouraged. Thank you so much, Dr. Ikseon Choi. I would like to recognize my committee members: Drs. Thomas C. Reeves, Lorilee Sandmann, and Michael Orey. Each of you provided unique and invaluable support. Dr. Reeves, thank you for your insightful guidance in crafting this document. You have a talent for editing and for helping others clearly understand key points! Dr. Sandmann, your suggestions on how to improve the literature review added clarity to the conceptual understanding of this study. Dr. Orey, I appreciate your insights on how to improve the design of this study.

A special note of appreciation goes to the research participants in my study. You know who you are. I appreciate your taking time out of your busy schedules and generously sharing your evaluation experiences with me, so that, together, we can try to improve how evaluations are conducted. Although this research is contextualized in the domain of evaluation of HIV/AIDS health education programs, your contributions will have much wider implications for how decision making is performed. A final thank you to my friends that encouraged me along the way: Rui Hu, Yi-Chun Hong, Elizabeth May, Romeo Santos, and Kwan-Chun Chen. Most of all, I thank God for His ever present help.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES x

LIST OF FIGURES xii

CHAPTER

1 INTRODUCTION 1

Background 5

Problem Statement 9

Research Focus 9

Research Questions 10

Importance of the Study 10

Glossary: Terms and Definitions 12

Overview of the Methodology 14

Chapter Summary 14

2 LITERATURE REVIEW 16

Overview of Evaluation Theories 16

Classical Decision Making Theories 20

Naturalistic Decision Making Framework 21

Evaluation as Naturalistic Decision Making: A Multiple-Case Study 34

Conceptual Frameworks Used in this Study 39

Guiding Framework for this Research 46

Chapter Summary 49

3 METHODOLOGY 51

Research Design 51

Research Participants 65

Data Gathering 66

Benefits and Risks 69

Limitations 70

Validation 72

Generalizability 73

Researcher's Perspective 74

Ethical Considerations 75

Pilot Study Findings 76

Chapter Summary 78

4 CASE DESCRIPTIONS 80

Case One: Formative Evaluation of a HIV Health Department and an Interview of a Department Official 85

Case Two: Summative Evaluation of an HIV/AIDS Health Education Program's Consortium Model 90

Case Three: Summative Impact Evaluation of a Micro-Credit and an HIV/AIDS Health Education Intervention 94

Case Four: Ex-Ante Evaluation That Was Aimed at Designing a Transactional Sex and Women's HIV/AIDS Vulnerability Evaluation 100

Case Five: Summative Evaluation of a Secondary-School HIV Health Education Intervention 109

Case Six: Design and Development of the Formative Evaluation of a Country's HIV/AIDS Home-Based Healthcare Supervision Programs 117

Case Seven: Formative Evaluation of the Health Education Component of a Surgical Intervention that Reduces HIV Transmission 125

Summary Overview of the Seven Cases 133

Chapter Summary 140

5 DATA COLLECTION, DATA REPRESENTATION, AND DATA ANALYSIS ..142

Chapter Purpose and Structure 142

Preliminary Research Activities 144

CTA's Three Phases, the Critical Decision Method, and Narrative Analysis 145

Phenomenography 178

Chapter Summary 179

6 RESULTS 180

Research Question One: What critical incidents do expert evaluators encounter during evaluations of HIV/AIDS health education programs? 180

Research Question Two: What factors influenced expert evaluators' naturalistic decision making during evaluations of HIV/AIDS health education programs? 201

Research Question Three: How do expert evaluators make naturalistic decisions when evaluating HIV/AIDS health education programs? 224

Variation of Critical Incidents, Factors, and Decision Making 240

Chapter Summary 243

7 FINDINGS, IMPLICATIONS, AND CONCLUSIONS 245

Research Design.....245

Summary of Findings.....246

Implications for Evaluation Practice.....269

Conclusions.....274

Implications for Future Research.....276

Chapter Summary 277

REFERENCES 278

APPENDICES

A PILOT STUDY 303

B IRB FORM.....335

C CONSENT FORM.....337

D DECISION MAKING TIMELINES AND MODELS 341

LIST OF TABLES

Page

Table 2.1: Empirical Research on Naturalistic Decision Making.....24

Table 3.1: Qualitative Philosophical Assumptions and How They Were Addressed 53

Table 3.2: Sweep Descriptions and Interview Questions 64

Table 4.1: Research Participants' Profiles.....134

Table 4.2: Evaluation Type and Evaluation Criteria 135

Table 4.3: Evaluation Type: External and Internal Evaluation 136

Table 4.4: Evaluation Designs and Data-Gathering Instruments 137

Table 4.5: Agencies that Commissioned Evaluations 137

Table 4.6: Summaries: Seven Case Narratives.....138

Table 5.1: Sweep One: Purpose and Interview Questions 147

Table 5.2: Sweep Two: Purpose and Interview Questions.....148

Table 5.3: Sweep Three: Purpose and Interview Questions 150

Table 5.4: Sweep Four: Purpose and Interview Questions.....152

Table 5.5: Sweep Five: Purpose and Interview Questions 155

Table 5.6: Sweep Six: Purpose and Interview Questions 157

Table 5.7: Case One: Critical Incident, Evaluator's Decisions, and Quotes 169

Table 5.8: Case One: Factors that Influenced the Evaluator's Decision Making.....170

Table 5.9: Case One Situation Assessment Based on Evaluation Resistance 174

Table 6.1: Barriers to Communicating with Stakeholders.....181

Table 6.2: Barriers and Enablers to Community Entry.....184

Table 6.3: Barriers to Understanding the Evaluand's Context 189

Table 6.4: Budget Constraints 192

Table 6.5: Evaluation Resistance.....193

Table 6.6: Threats to High Quality Data 195

Table 6.7: Threats to the Evaluation Design 196

Table 6.8: Time Constraints 198

Table 6.9: Critical Incidents 201

Table 6.10: Context Factors 206

Table 6.11: Evaluation Factors 213

Table 6.12: Human Factors.....218

Table 6.13: Real-World Factors.....223

Table 6.14: Case Four Situation: Assessment Understanding the Evaluand's Context 229

Table 6.15: Critical Incidents: Barriers or Enablers, and Expected or Unexpected 240

Table 7.1: Real-World Factors Categories, Types, and Definitions.....259

Table A1.1: Critical Incident-Decision Relationship with Corresponding Quotes 311

LIST OF FIGURES

Page

Figure 2.1: Recognition-Primed Decision Making Model 41

Figure 2.2: Conceptual Framework for How Evaluators Make Practice Decisions 43

Figure 2.3: Planning a evaluation of a development program 47

Figure 2.4: Conceptual framework of program evaluators' naturalistic decision making process during critical incidents 49

Figure 3.1: Cognitive Task Analysis' three phases and the Critical Decision Method 55

Figure 3.2: Research design and procedure 59

Figure 3.3: Narrative analysis procedure 63

Figure 3.4: Pilot study: Expert evaluator's naturalistic decision making model 78

Figure 5.1: Case One: Rough draft decision making timeline with interview questions 161

Figure 5.2: Case One: Finalized evaluation resistance decision making timeline.....162

Figure 5.3: Case One memo: Politics, evaluator's frustration, and researcher's reflections 167

Figure 5.4: Case One memo: Impact of politics on stakeholders, occurring in the phenomenon under study 168

Figure 5.5: Initial stand-alone data analysis 168

Figure 5.6: Case One Model: Evaluation Context, Critical Incident, Situation Assessment, and Decision Making Process.....176

Figure 7.1: Expert evaluators' naturalistic decision making during critical incidents 262

Figure 7.2: Expert Evaluator's Decision Making Cycle.....264

Figure 7.3: Variation among the critical incidents that the evaluators encountered, factors that influenced evaluators' naturalistic decision making, and the methods used when evaluators made decisions.....267

Figure A1.1: Evaluator's critical incidents, decisions, and timeline 310

Figure A1.2: Detailed program evaluator's decision model.....310

Figure A1.3: High-level overview of program evaluator's decision making process 311

Figure A2.1: Narrative based on Pilot Study: A Program Evaluator's Decision Making When Facing Critical Incidents.....328

Figure A3.1: Pilot study interview protocol and questions 330

Figure A4.1: Pilot study consent form.....334

Figure D1.1: Case 1 Decision Making Timeline - Formative evaluation of an HIV health department and an interview of an evaluation department official 341

Figure D1.2: Case 1 Model - Formative evaluation of an HIV health department and an interview of an evaluation department official 342

Figure D1.3: Case 2 Decision Making Timeline - Summative impact evaluation of an HIV/AIDS health education program's consortium model 343

Figure D1.4: Case 2 Model - Summative impact evaluation of an HIV/AIDS health education program's consortium model ……..344

Figure D1.5: Case 3 Decision Making Timeline - Summative impact evaluation of micro-credit and an HIV/AIDS health education intervention 345

Figure D1.6: Case 3 Model - Summative impact evaluation of micro-credit and an HIV/AIDS health education intervention.....346

Figure D1.7: Case 4 Decision Making Timeline - Ex-Ante evaluation that was aimed at designing a transactional sex and women's HIV/AIDS vulnerability evaluation 347

Figure D1.8: Case 4 Model - Ex-Ante evaluation that was aimed at designing a transactional sex and women's HIV/AIDS vulnerability evaluation …...……..348

Figure D1.9: Case 5 Decision Making Timeline - Summative evaluation of a secondary school HIV health education intervention 349

Figure D1.10: Case 5 Model - Summative evaluation of a secondary school HIV health education intervention 350

Figure D1.11: Case 6 Decision Making Timeline - Design and development of the formative evaluation of a country's HIV/AIDS home-based healthcare supervision programs.....351

Figure D1.12: Case 6 Model - Design and development of the formative evaluation of a country's HIV/AIDS home-based healthcare supervision programs.....352

Figure D1.13: Case 7 Decision Making Timeline - Formative evaluation of the health education component of a surgical intervention that reduces HIV transmission 353

Figure D1.14: Case 7 Model - Formative evaluation of the health education component of a surgical intervention that reduces HIV transmission.....354

CHAPTER 1

INTRODUCTION

The purpose of this study is to understand how expert evaluators make naturalistic decisions in the context of evaluating HIV/AIDS health education programs. The results of this investigation are intended to assist other real world decision makers, in general, to become more informed about the factors that shape real world decision making and to assist HIV/AIDS evaluators, in particular, to understand the critical incidents expert evaluators encountered and the factors that influenced their decision making. This study also has implications for the design of new curricula that attempts to increase evaluation capacity building. Specifically, the findings from the narratives, decision making timelines, and models derived from interviews with expert evaluators will be incorporated in the design of these new curricula. It is expected that increased evaluation capacity will improve the evaluation of HIV/AIDS health education programs, thereby positively impacting the fight against HIV/AIDS.

HIV/AIDS health education evaluators are working to gather credible evidence to determine whether international aid agencies' HIV/AIDS health education programs are effective, relevant, sustainable, or meeting their objectives (Mantell, DiVittis, & Auerbach, 1997). Many theories and models are available that guide evaluators as they conduct evaluations (Stufflebeam & Shinkfield, 2007). However, there is limited knowledge on how expert evaluators make naturalistic decisions (Kundin, 2008, 2010; Tourmen, 2009); also, referred to as how people make decisions in a real-world context (Zsambok & Klein, 1997). A review of the existing literature on evaluation practice indicated that there is a lack of research on how expert

evaluators make naturalistic decisions when conducting evaluations of HIV/AIDS health education programs. This study has been designed to increase the understanding of how evaluators make decisions when conducting evaluations of HIV/AIDS health education programs, thus filling an important gap in the literature.

Evaluators often make decisions either in the field or "along the way" (Stufflebeam & Shinkfield, 2007, p. 20), because "often one [the evaluator] cannot precisely define beforehand the appropriate standards, evaluative criteria, and defensible levels of soundness for each one and for all as a group" (p. 20). Stufflebeam and Shinkfield have recommended that an interactive discourse between evaluators and stakeholders guide decision making throughout "the evaluation design, analysis, and interpretation" (p. 20). However, there is limited research that provides an insight into how evaluators and stakeholders communicate with each other (Hogard & Ellis, 2006) and how evaluators make decisions (Kundin, 2008, 2010; Tourmen, 2009).

When conducting evaluations, evaluators frequently encounter critical incidents (Kundin, 2008, 2010). For example, critical incidents can occur, because evaluators frequently need to "address competing and often conflicting values of different members of an evaluation audience" (Stufflebeam & Shinkfield, 2007, p. 21). Furthermore, evaluators must also "work within his or her own philosophical convictions," (Stufflebeam & Shinkfield, 2007, p. 22) and their decisions are inevitably influenced by their "positionality" and "values" (Mertens & Wilson, 2012, p. 165). A review of evaluation literature found that there is limited research available on the types of critical incidents that HIV/AIDS health education evaluators encountered and also on the factors that influenced their naturalistic decision making when they planned, conducted, and reported evaluations.

In this study, *expert evaluators* are defined as evaluators with at least five years of evaluation experience. Ericsson, Krampe, and Tesch-Römer's (1993) research indicates that an expert is a person with at least 10,000 hours of deliberate practice. Based on a five-day, 40-hour work week in a 52-week year, it was estimated that 10,000 hours of deliberate evaluation practice is equal to five years of evaluation experience. Although it would be interesting to understand how novice evaluators make decisions in the field, I have focused on expert evaluators since they tend to have more evaluation experience, usually deal with more critical incidents, and, therefore, have more stories to share (Schank, 1990).

The evaluation of HIV/AIDS health education programs implemented in SSA was chosen as the research context for this study. I chose to study evaluator's naturalistic decision making when evaluating HIV/AIDS health education programs because of the importance of these programs to the people in Sub-Saharan Africa (SSA). In addition, these types of educational programs are relied on in order to increase knowledge about HIV/AIDS and inform people on how to avoid or reduce the risk of an HIV/AIDS infection (Baker, Leon, & Collins, 2011). HIV/AIDS health education programs work under the premise that a "heightened awareness and knowledge of health risks are important preconditions for self-directed change" (Bandura, 1990, p. 9). These programs are considered a "social vaccine" that "lower[s] unhealthy risk-taking and lead[s] to an effective preventative" (Baker, Leon & Collins, 2011, p. 1319). Another reason for investigating evaluators' decision making within the context of HIV/AIDS health education programs implemented in SSA is that the population in SSA has been disproportionately impacted by this serious pandemic (UNAIDS, 2008). Lessons learned in this challenging area should have implications for HIV/AIDS health education programs around the globe.

In response to the devastating impact of HIV/AIDS in developing countries, international aid agencies have been working on many fronts. The actors in global health assistance are complex, including initiatives from "philanthropic trusts and other civil society organizations, private-for-profit organizations, global health initiatives and partnerships" (Spicer & Walsh, 2010, p. 1). Many of these global health initiatives involve HIV/AIDS education programs that are intended to address the complex issues surrounding HIV/AIDS, but research indicates that health education programs have had mixed outcomes in changing behaviors which reduce the risk of contracting HIV/AIDS (Fisher, 1992). For example, in a study of an HIV/AIDS health education program for in-school adolescents in Nigeria, researchers found that students' knowledge about HIV/AIDS did not "significantly influence" their risky "sexual behavior" (Okonata & Oseji, 2006, p. 37). A literature review by Auerbach, Hayes, and Kandathil (2006) included an analysis of the impact of twenty-two evaluations of sex education and HIV education interventions implemented in schools in developing countries that were published between 1990 and 2005. The study found that the majority of the programs implemented in developing countries "demonstrated that their interventions produced positive behavioural changes that, logically, should lead to reductions in STIs [Sexually Transmitted Infections] and HIV" (Kirby, Obasi, & Laris, 2006, p. 145), but evidence that STIs and HIV had actually declined was not found. Many of these interventions, as well as other attempts to reduce or eliminate the spread of HIV, have been "tested and evaluated … [but] relatively few have been evaluated systematically and rigorously" (Auerbach et al., 2006, p. 43).

Unfortunately, it is evident that international aid agencies health education programs which are designed to halt the Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) pandemic have had "less than optimal results" (Campbell,

2003, p. 18). Although the infection rate in some countries has been reduced, other countries have seen an increased rate of HIV/AIDS transmission. One example of this reduction is a UNAIDS (2011) report stating that "The number of people dying of AIDS-related causes fell to 1.8 million [1.6 million–1.9 million] in 2010, down from a peak of 2.2 million [2.1 million–2.5 million] in the mid-2000s." It also reported that "an estimated 34 Million people [31.6- 35.2 million] live with AIDS worldwide … [and] Sub-Saharan Africa women are more affected … (59% of all people living with HIV)" (p. 6). At the same time, although these data may indicate that the transmission of HIV/AIDS has been reduced, a significant number of people are still at the risk of being infected. For example, from 2001 to 2010, there was a 250% increase in the number of people living with HIV/AIDS in Eastern Europe and Central Asia (UNAIDS, 2011).

Inconsistent reports regarding the effectiveness of HIV/AIDS health education programs highlight the need to critically assess the "effectiveness of the substantial investments by bi- and multi-lateral agencies, NGOs and governments in the regions [SSA] whose prevention programs chiefly target the acquisition of facts and positive attitude" (Baker, Leon, & Collins, 2011, p. 1352). There is confusion as to whether HIV public health interventions have been effective (Ross, Wight, Dowsett, Buve, & Obasi, 2006), and researchers have indicated that this may be because the evidence used to "make decisions about interventions to prevent the spread of HIV is extremely complex … and arising from diverse evaluation methods" (Ross et al., 2006, p. 99).

Background

This section provides a brief background on evaluation, development, and the naturalistic decision making research. *Evaluation* is defined as "an assessment, as systematic and objective as possible, of a planned, on-going, or completed intervention" (OECD, 2010, p. 21) to determine "the worth or merit" (Joint Committee on Standards of Education Evaluation, 1994, p.

3) of an intervention. When conducting evaluations, there is a "need to choose the appropriate values of judging an evaluation" (Stufflebeam & Shinkfield, 2007, p. 9). Evaluations are usually classified as either formative or summative (Scriven, 1967). Formative and summative evaluations are conducted at different phases in the lifecycle of a program and have different goals (Fitzpatrick, Sanders, & Worthen, 2011). Formative evaluations are usually conducted during the design and early implementation phases to improve a program or maintain quality control (Stufflebeam & Shinkfield, 2007).

"Evaluation practice refers to the definition of the evaluation, the research design, and the execution of the evaluation activity, that is, implementation, results, and impacts on specific public policy" (Boyle, Lemaire, & Rist, 1999, p. 5). Some researchers have traced evaluation's history either to Socrates' question-oriented interactions with his students (Fitzpatrick et al., 2011) or to the Old Testament's book of Daniel, which includes a description of King Nebuchadnezzar's instituting an evaluation of a civil service educational program (Patton, 2008). Recently, the administration of U. S. President Barack Obama in an effort to meet the country's educational goals emphasized the use of evaluation to improve programs (Office of Management and Budget, 2009).

Program evaluations rely on multi-disciplinary approaches (Patton, 2008). Program evaluation practice is influenced by auditing (Wisler, 1996) as well by other scientific and social science traditions (Rossi & Freeman, 2004). As program evaluations became commonplace in the wake of large U. S. federal government investments in education and other social programs, development evaluation emerged as a distinct focus for program evaluators. According to Earl and Carden (2002),

> Development is a complex process comprising three parallel dynamics: first, the changes in the behaviours, actions, and/or relationship of the people, groups, and

> organisations with whom a programme works directly; second, the strategies that a programme employs to encourage change in its partners; and, third, the functioning of a programme as an organizational unit. (p. 520)

Development evaluations are conducted with the aim of increasing the understanding of international development aid agencies programs' "sustainability, effectiveness, efficiency, impact, relevance, and fulfillment of objectives" (OECD, 2010, p. 20). The majority of the knowledge about international development programs is generated from the evaluations performed during different stages in a program's life cycle (Morra Imas & Rist, 2009). Many international development programs are supported by and carried out through joint international partnerships (Solomon & Chowdhury, 2003). These partnerships are used to address global issues, such as the millennium development goals (MDGs). The MDG 6.A has as its goal to "halt and reverse the spread of the HIV" (UNAIDS, 2010, p. 7) epidemic, an epidemic in which SSA "bears an inordinate share of the global HIV burden" (p. 25).

As stated by the United Nations Educational Scientific and Cultural Organization (UNESCO), "there is no universal path of sustainable development" (Becker, Jahn, Stiess, & Wehling, 1997, p.20). Due to a need for novel international development approaches (Regeer, Hoes, van Amstel-van Saane, Caron-Flinterman, & Bunders, 2009), evaluators have develop new evaluation approaches as well. Consequently, international development evaluations have become more complex, requiring evaluators to use tools from multiple disciplines as well as a systems approach to deal with the many different variables typically interacting within development programs (Chelimsky & Shadish, 1997; Patton, 2006). There has also been a need to create novel evaluation architectures (Mantell, DiVittis, & Auerbach, 1997; Patton, 2006) as international aid entities try to develop and implement initiatives on a global scale.

As evaluators navigate through an evaluation's uniqueness, context, complexity, and multi-disciplinary components, "evaluators bring to the negotiating table their own style, personal history and professional experience" (Patton, 2008, p. 214). How challenges and opportunities present themselves in an evaluation, as well as the characteristics that evaluators bring to an evaluation, are not usually included in evaluation reports (Christi, 2009). In addition, the decisions that evaluators make when conducting evaluations have not been extensively researched (Christie & Rose, 2003; Kundin, 2008, 2010).

Classical decision making research differs from naturalistic decision making research (Zsambok & Klein, 1997). Classical decision making research is usually conducted to study novices in a laboratory in which the decisions are made "apart from any meaningful context" (Orasanu & Connelly, p. 8). I determined that studying decision making in a context-free laboratory setting would limit my ability to understand how actual expert evaluators make decisions. Since evaluators often require contextually based "situational responsiveness" (Patton, 2008, p. 28), I decided that it was important to study program evaluation in its naturalistic setting rather than in a context-free laboratory environment.

Naturalistic decision making characteristics are difficult to control. It is, therefore, not always possible to observe or interview participants as they conduct tasks. As a result, elicitation techniques have been developed. One of these techniques is the Critical Decision Method. The critical decision method consists of conducting retrospective interviews in which experts recall their decision making when they are confronted with critical incidents (Crandall, Klein, & Hoffman, 2006).

It can be difficult to interview or study an evaluator either during or immediately after completing an evaluation. Therefore, I decided to use retrospective interviews based on the

critical decision method in my study. As just noted, the main purpose of this study is to understand how expert evaluators make naturalistic decisions when they are evaluating HIV/AIDS health education programs.

Problem Statement

The HIV/AIDS epidemic has had significantly devastating effects in SSA (Campbell, 2003), and international aid agencies are implementing HIV/AIDS health education programs to stem the spread of HIV/AIDS. Their efforts have had mixed results. Evaluations of HIV/AIDS health education programs are conducted to inform current programs and assist in the development of future programs that are utilized to stem the epidemic. However, research on how evaluators conduct evaluations and on evaluators' decision making is limited (Kundin, 2008, 2010; Tourmen, 2009). Furthermore, a review of the evaluation literature found that research focusing on expert HIV/AIDS health education program evaluators' naturalistic decision making has not been conducted.

Research Focus

The specific purpose of this study was to understand how evaluators make naturalistic decisions in the aforementioned research context. To inform my study, I decided to research the critical incidents that evaluators faced and the factors, which influenced their decisions to develop an understanding of how they make naturalistic decisions. Co-constructed member checked narratives generated through retrospective semi-structured interviews were utilized to identify the critical incidents that evaluators encountered and the factors that influenced their decisions, and to understand how expert evaluators make naturalistic decisions. The narratives will be developed into cases that will be used to develop future curricula that will provide novice

and experience evaluators with an opportunity to enhance their understanding of evaluation practice.

Research Questions

This study utilized the Critical Decision Method (CDM), a cognitive task analysis method. CDM was utilized to study expert evaluators' naturalistic decision making when they encountered non-routine situations that required expert decisions or judgment as they evaluated HIV/AIDS health education programs. The critical incidents that evaluators encountered were analyzed to determine the factors that influenced their naturalistic decision making. The experts interviewed were evaluators who have evaluated HIV/AIDS health education programs.

The following questions were addressed in this study:

1. What types of critical incidents do expert evaluators encounter during evaluations of HIV/AIDS health education programs?
2. What factors influence expert evaluators' naturalistic decision making during evaluations of HIV/AIDS health education programs?
3. How do expert evaluators make naturalistic decisions during evaluations of HIV/AIDS health education programs?

Importance of the Study

According to a 2008 report on the Global AIDS Epidemic, the Joint United Nations Program on HIV/AIDS (UNAIDS) stated that in 2007, there were "33 million people living with HIV/AIDS" worldwide (p. 5). The report also stated that "Sub-Saharan Africa accounts for 67 percent of the people living with HIV and for 72 percent of AIDS deaths" (p. 5). As previously noted, international development agencies are funding HIV/AIDS health education programs in SSA to reduce the impact of this disease. Their actions are consistent with Sattler and Armmer's

(2006) statement, that "Education is the primary tool we have today to combat the world's HIV/AIDS pandemic" (p. 75).

Nyanzi (2006) warned:

> Nearly three decades of prevention interventions against HIV/AIDS have yielded little effect. . . . What is missing in available sexual-health programs, policies, and activism? Why are they not as effective as they promise? What is wrong with these interventions? (p. 1851)

HIV/AIDS prevention interventions are relying on evaluations of their programs to answer Nyanzi's challenging questions. Evaluations are an essential component of the iterative process of understanding, executing, and fine tuning, providing quality control, and determining the impact of an intervention (Stufflebeam & Shinkfield, 2007), and, hence, any effort to enhance evaluation practice is important. A 2008 UNAIDS report highlighted that organizations should base their interventions on sound evidence. In that vein, stakeholders of HIV/AIDS health education programs are relying on evaluations to inform current programs by providing sound evidence and to assist in the development of future programs. They should also look to the results of evaluations to inform day-to-day decision making and quality control. Studying HIV/AIDS health education program evaluators' naturalistic decision making will inform the limited body of knowledge regarding evaluation practices, improve evaluation practices, and, as a result, improve programs.

In order to understand evaluator's naturalistic decision making, narratives depicting how evaluators conduct evaluations were gathered through retrospective interviews. I will eventually convert the expert evaluators' narratives into case studies that will be used to teach evaluation concepts and provide an opportunity for students to study cases based on authentic evaluation experiences. The cases will teach basic program evaluation concepts and depict the decisions

that evaluators make in the field. The cases will be utilized (a) for training and professional development, (b) to inform evaluation theories, and (c) to inform evaluation policies.

Glossary: Terms and Definitions

Cognitive Task Analysis (CTA) refers to a group of scientific approaches whose goal is to facilitate understanding of the thought processes of capable practitioners' reasoning and strategies in a complex work environment. Although considered a scientific approach, it is also an art form derived from multiple disciplines, especially applied and academic psychology. I used this definition, because it defines cognitive task analysis as a scientific and art-based approach that people in the CTA field use to obtain knowledge about how people think and the strategies they use to complete tasks as they navigate a changing and contextually defined setting.

Critical Incident (CI) refers to an event that is especially challenging and in which the performer's skills and knowledge are called on to address one or more issues. The events do not need to be rare, but they tend to be atypical. In this study, this definition is utilized to emphasize that the events do not need to be life threatening but rather require the use of an evaluator's knowledge or skills to address a situation encountered when conducting an evaluation. The events studied had an influence on the manner in which the evaluation was conducted or the evaluation's outcome.

Expert is defined as an individual with a high level of knowledge or ability in his or her domain. *Expert* and *expertise* are debated terms. Several researchers consider the terms *expert* and *expertise* as socially defined constructs and are, therefore, difficult to be measured. Sometimes, expertise is defined by a community of practice. Expertise in this study is defined based on research by Ericsson, Krampe, and Tesch-Römer's (1993) research, indicating that

10,000 hours of deliberate practice is generally necessary to obtain expertise. Based on this information, it was estimated that a full-time evaluator who worked at least 40 hours per 5-day workweek for 5 years would reach the 10,000-hour mark.

Health not only refers to being free of disease but also includes a person's physical, psychological, and social well-being. This definition is derived from the World Health Organization's definition of health (1946). Some development programs specialize in different states of health, as listed in the World Health Organization's definition.

Ill-structured Problems are also referred to as *wicked problems*. *Ill-structured problems* often have goals that are not easy ascertained or have missing components, which are needed for problem solving. The goals are ever changing as well as conflicting. They can occur at the strategic, project, and individual level. This definition was utilized, because it captures some of the characteristics of naturalistic decision making.

Naturalistic Decision Making (NDM) studies how experts make decisions as they endeavor to successfully complete tasks while working within in a group or as individuals in the setting where the tasks are being performed. This definition was utilized in the study, because it highlights that naturalistic decision making research studies experts' decision making in real-world settings.

Situation Assessment refers to the awareness of the state of an environment through the perceiving of cues, assigning meaning to cues, and utilizing the found meaning to predict their relevance to future events. I used this definition, because it takes into account the fact that a situation assessment is not only more than an awareness of what is occurring in an environment but also the process of gathering information to anticipate what could occur next.

Stakeholder refers to any group or individual or organizations that can profit, influence, or be influenced by the actions related to reaching a program's objective. I used this definition to capture the fact that the evaluator, program donors, and intervention beneficiaries as well as evaluation participants, are stakeholders. The stakeholders can be either directly or indirectly involved in or impacted by a program.

Overview of the Methodology

Qualitative methods best suited the exploratory nature (Creswell, 2009) of this research and the researcher's goal of understanding expert evaluators' naturalistic decision making when evaluating HIV/AIDS health education programs. Collective case study techniques (Stake, 1995) and phenomenographic methods (Marton, 1994) were used in this research, because they are helpful in investigating exploratory or little-researched phenomena. Data from seven interviews of expert evaluators were used in my study. Data collection continued until data saturation (Corbin & Strauss, 2008) was reached.

Chapter Summary

The goal of this research was to understand expert evaluators' naturalistic decision making during evaluations of HIV/AIDS health education programs. The focus was to understand the following: (1) the critical incidents that evaluators encountered; (2) the factors that influenced evaluators' naturalistic decision making; and (3) how evaluators made decisions. In order to understand evaluator's naturalistic decision making, narratives depicting how evaluators conduct evaluations were gathered through retrospective interviews. A study of HIV/AIDS health education program evaluators' naturalistic decision making will inform the limited body of knowledge regarding evaluation practices, improve evaluation practices, and, as a result, improve programs. This study also has implications for the design of new curricula that

attempting to increase evaluation capacity building. Specifically, the findings from the narratives, decision making timelines, and models derived from interviews with expert evaluators will be incorporated in the design of these new curricula.

An adaption of Cognitive Task Analysis and the Critical Decision Method (Crandall, Klein, & Hoffman, 2006) guided my study. These two approaches assisted in knowledge elicitation, knowledge representation, and data analysis. I used inductive data analysis in this study, and employed open coding (identification of themes, codes, and criteria) to understand the critical incidents that evaluators faced and the decisions they made, and to identify the factors that influenced the evaluators' decision making. The methodology and research procedures are thoroughly explained in Chapter Three.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a literature review of the relevant theoretical, conceptual, and practical research used to inform this study. In particular, the first section provides an overview of the evaluation theories used to establish a foundation for my research. The second section reviews classical decision making and why it is not an appropriate decision making approach for this study. The third section examines naturalistic decision making in detail, as it sets the framework for the decision making model used in this study (Fitzpatrick, Christi, & Mark, 2004). The fourth section provides an overview of a multiple-case study conducted on naturalistic decision making and evaluation to illustrate how key factors in naturalistic decision making are played out in real world evaluations. The fifth section details Klein's Recognition-Primed Decision Model (1997) and Kundin's Conceptual Framework for How Evaluators Make Practice Decisions (2008, 2010), which are exemplars of naturalistic decision making models and, thereby, shapes the guiding conceptual framework for this study. The final section develops the conceptual framework of this study.

Overview of Evaluation Theories

According to Branch (2009), evaluation "is a distinct discipline worthy of in-depth study about its theories and practices" (p. 153). Ideally, evaluations are conducted throughout the "life cycle of an initiative" (Kusek, & Rist, 2004, p. 3). Different types of evaluations occur during different stages in a program's life cycle and have different purposes (Stufflebeam & Shinkfield, 2007). Formative evaluations are usually conducted while implementing or improving programs

(Stufflebeam & Shinkfield, 2007). Summative evaluations are conducted after program implementation to ascertain program effectiveness (i.e., whether a program achieved its expected outcomes; Stufflebeam and Shinkfield, 2007).

Although the definitions of and distinctions between formative and summative evaluations appear straightforward, the real world implementation of these two types of evaluations is not as straightforward (Patton, 2008, 2010). Evaluators' philosophical assumptions underlying formative and summative evaluations color how evaluators define formative and summative evaluations in practice and how they conduct these evaluations. Mertens (2008) lists four major philosophical assumptions that shape evaluators' conceptualizations and implementations of evaluation: (1) axiological, (2) ontological, (3) epistemological, and (4) methodological. This results in conflicting definitions of "evaluation" (Alkin, 2004; Alkin & Christi, 2004) and numerous "alternative … [if not] incompatible, paradigms and diverse models" (Reeves & Hedberg, 2003, p. 30). These paradigms include the following types of evaluation paradigms and models: Analytic-Empirical-Quantitative, Constructivist-Hermeneutic-Interpretivist-Qualitative, Critical Theory-Neomarxist-Postmodern-Praxis, and Eclectic-Mixed Method-Pragmatic paradigms (Reeves & Hedberg, 2003).

In addition to differing philosophical assumptions, other factors also shape evaluators' understanding and implementation of evaluations. Christi and Azzam (2005) report that there is a "diversity of theoretical perspectives presented" by different evaluation theorists (p. 15). For example, evaluators' perspectives regarding experimental versus non-experimental designs as well as qualitative, quantitative, and mixed-method research approaches (Mark, Henry & Julnes, 1999; Mertens, 2008) shape the theoretical frameworks that define what is appropriate evaluation

research. Donaldson, Christi, and Mark (2008) add that varying theoretical perspectives determine what evaluators regard as trustworthy evidence.

The contexts in which evaluations are conducted further add to the diversity of evaluation understanding and practice. For example, evaluations are conducted in education where the focus is to assess the efficacy of a learning approach, among other factors. Evaluations are also routinely conducted to assess the efficacy of health education programs. It is the goal of this research to focus on evaluations in the context of international development health education programs. This context introduces a number of other factors that also shape how evaluations are conducted.

One of these factors concerns the evaluator's role in the organization or program being evaluated. Specifically, is the evaluator internal or external to the program or organization being evaluated? What role does the evaluator play? According to Fitzpatrick, Sanders, and Worthten (2011), the role of the evaluator is evolving from expert to teacher to facilitator to political and to advisory roles.

Increasingly, stakeholder participation is also coloring how evaluations are conducted. Rossi, Lipsey, and Freeman (2004) report that, in recent years, stakeholders have become more involved in the evaluation process and, therefore, are advocating what should be evaluated and how it should be evaluated. This has raised awareness that we need to conduct more culturally/contextually responsive evaluations (Hood, Hopson, & Frierson, 2005) as well as to use evaluations to address social justice issues (Mertens, 2008). There have also been discussions on how to create a culture of evaluation; that is, to build evaluation capacity (Compton, Baizerman, & Stockdill, 2002).

Although evaluations and their implementations differ for a number of reasons, they do share some common themes and seem to be evolving toward more complex activities. According to the Organization for Economic Co-operation and Development (OECD, 1991), "The aim of development program evaluation is to: (a) determine the program's relevance, (b) fulfillment of objectives, (c) efficiency, (d) effectiveness, (e) impact and (f) sustainability" (p. 4). Evaluations are evolving from traditional implementations that focus "on evaluation of inputs, activities and outputs (that is, on project or program implementation)" (Morra Imas & Rist, 2009, p. 108) toward hybrid approaches. Results-Based Monitoring and Evaluation represents such an approach. It "combines the traditional approach of monitoring and implementation with assessment of outcomes and impacts, or more generally of results" (Morra-Imas & Rist, 2009, p. 108). Fitzpatrick, Sanders, and Worthen (2011) offer practical guidelines for planning evaluations from which we can deduce questions that can guide an evaluation. These guiding questions assist an evaluator to "determine what should be evaluated, how to be certain the evaluation is focused on the right things and how to plan the specifics of an evaluation" (Fitzpatrick et al., 2011, p. 379):

(1) What is the purpose of the program evaluation?

(2) What aspects of the program are being evaluated?

(3) Who are the evaluation and program stakeholders?

(4) What is the environment in which the evaluation will take place?

(5) What and which evaluation questions should be answered?

(6) What and which criteria should be used?

(7) How and when will information be needed, collected, analyzed, and interpreted?

(8) What resources will be needed to collect, analyze, and interpret the information?

(9) How will the findings be reported and to whom?

(10) How will the evaluation be used?

Classical Decision Making Theories

Classical decision making emerged from economists' game theory research, which considers human beings as rational actors who make rational decisions (von Neumann & Morgenstern, 1947). In the early 1950s, Edwards (1954) conceived of a variation of the classical decision theory in which he proposes that human beings make rational decisions based on subjective rather than objective factors. Edwards (1954) found, and Svenson and Maule (1993) later confirmed, that individuals do not use objective values to critique the usefulness of outcomes. Instead, they critique the usefulness of outcomes based on subjective probability and subjective utility values. This research led Edwards to propose the Subjective Expected Utility Approach.

Beach (1990) extended Edwards' (1954) ground-breaking work by proposing the Multi-Attribute Utility Analysis approach to decision making. Beach found that previous studies did not capture the complexity associated with decision making, because they did not consider the multiple factors that shape decision making in relation to each other. Kahneman, Slovic, and Tversky's (1982) research expanded Beach's complex decision making by incorporating account biases and heuristics to simplify decision making. Their approach utilizes the Bayesian decision theory and introduces the concept that decision makers use heuristics to reduce the complexity of the decision making process.

What is distinctive about classical theoretical decision making approaches is that decision making is typically conducted in laboratory settings and uses college students (Zsambok & Klein, 1997). Although there is merit in each of these classical theoretical approaches, they do

not address how decisions are made in the field by experts. For these reasons, classical decision making is not appropriate for this study. Naturalistic decision making research aims to fill these gaps.

Naturalistic Decision Making Framework

Naturalistic decision making (NDM) introduces two important factors to research on decision making: (1) the real world context and (2) real world experts (Zsambok & Klein, 1997). In doing so, NDM takes a contrasting position to classical decision making. NDM researchers "have been motivated by curiosity about how people do so well under difficult situations" (Klein, 1998, p. 1). This requires that we understand "how experienced people, working as individuals or groups in dynamic, uncertain, and often fast-paced environments, identify and assess their situation, make decisions and take actions whose consequences are meaningful to them and to the larger organization" (Zsambok & Klein, 1997, p. 5). NDM, therefore, relies on studying contextual factors and the cognitive processes that guide how experts make decisions.

During the 1989 NDM conference, researchers identified key contextual factors that distinguish NDM from classical decision making (Zsambok & Klein, 1997, p. 5):

(1) Ill-structured problems (not artificial, well-structured problems).

(2) Uncertain, dynamic environments (not static, simulated situations).

(3) Shifting, Ill-defined, or competing goals (not clear and stable goals).

(4) Action/feedback loops (not one-shot decisions).

(5) Time stress (as opposed to ample time for tasks).

(6) High stakes (not situations devoid of true consequences for the decision maker).

(7) Multiple players (as opposed to individual decision making).

(8) Organizational goals and norms (as opposed to decision making in a vacuum).

Naturalistic (1997) research has found that decision makers rely on satisficing (Simon, 1957, 1991) when making decisions. Simon found that the heuristic aspect of decision making introduces the concepts of maximizing and satisficing; that is, making decisions based on the concepts that they are either optimal or satisfy most of our requirements. According to Campitelli and Gobet (2010),

> Simon's approach to decision making essentially consisted of three main assumptions: first, decisions are not performed by agents with perfect rationality, they are made by agents with bounded rationality; second the quality of decisions vary as a function of expertise of the decision maker; third to understand decision making, it is paramount to investigate the cognitive processes involved; that is, an analysis based on performance only is not sufficient. (p. 354)

Simon's research identified two cognitive decision making styles: maximizing—making an optimal decision (where no other decision can have a better outcome), and satisficing—selecting a decision that meets a minimum-threshold or aspiration level based on a criteria or a degree of satisfaction. Simon (1991) concluded that our decisions are confined by a "bounded rationality" (p. 125) that is shaped by our minds' abilities to process information and by the amount and quality of information available (which is usually limited). Satisficing occurs due to the complexity of "the human environment and the limitations of human processing …. To satisfice is to pursue not the best option but a good enough option" (Schwartz, Ward, Monterosso, Lyubomirsky, & Lehman, 2002, p. 1178).

Since real world decision making is very complex and incorporates many different factors and sources of information (Bamberger, Rugh, & Mabry, 2006; Patton, 2010), I decided to incorporate Simon's approach to decision making with an emphasis on satisficing.

After the 1994 Naturalistic Decision Making Conference, Zsambok and Klein (1997) edited a compilation of articles from researchers and practitioners that utilized naturalistic decision making to understand experts' decision making in disciplines such as medicine (Boger,

1997), the military (Serfaty, MacMillan, Entin, & Entin, 1997), industry (Roth, 1997), and computer science (Mitchell, Morris, Ockerman, and Potter, 1997). The evaluators also studied how experts worked in teams as well as during training (Salas, Cannon-Bowers, & Johnston, 1997).

Table 2.1 provides examples of recent (2009–2012) empirical research that complements naturalistic research. The articles exemplify blended approaches in naturalistic research. Some researchers combine CTA and Cognitive Work Analysis (Jenkins et al., 2010) to shed light on how participants made naturalistic decisions; others illustrate the use of different data gathering techniques; and still others use simulations to capture experts' field experiences (Boyes & O'Hare, 2011; Chavin et al, 2009; Helsloot & Groenendaal, 2011; Jenkins et al., 2010; Marshall et al, 2011), observations (Baber, Fulthrope, & Houghton, 2010; Franklin, et al., 2011; Marshall et al., 2011;Vajaramta & Easton, 2010), and interviews (Baretto & Riberio, 2012; Helsloot & Groenendaal, 2011; Jenkins et al., 2010).

Interview techniques such as observations with think-aloud protocols (Franklin et al, 2011) and observations have also been combined with computer simulations (Boyes and Hare, 2011; Helsloot & Groenendaal, 2011; Marshall et al., 2011). The articles also reveal alternative techniques used for observing participants' field experiences, including indirect observations by studying the recordings of a business meeting (Vajaramta & Easton, 2010) or studying communications among power plant operators (Greitzer, Podmore, Robinson, & Ey, 2010).

Table 2.1

Empirical Research on Naturalistic Decision Making

Study	Purpose	Method	Relevant Findings	Challenges
Baber, Fulthorpe, & Houghton (2010)	To explore the possible use of location-based photography for field-based intelligence gathering.	Team patrol members were randomly assigned to record observations on paper and a map. They were given a location-based imaging camera to aid their observations. Reports were analyzed using context analysis.	Participants, using location-based imaging prototypes, captured important features in the environment and were able to provide well-reasoned conclusions.	Patrols were not observed while patrolling the situation.
Baretto, & Ribeiro (2012)	To explore whether real world factors influenced aeronautical investigators' accident investigations and whether they made naturalistic decisions.	Used the Critical Decision Method to interview seven aircraft accident investigators with over two years of experience.	Aircraft accident investigators used naturalistic decision making strategies such as mental simulation, intuition, and improvisation.	Interviews were used which precluded the researchers from observing the investigators in the field.

Boyes, & O'Hare (2011)	To examine the NDM process followed by leaders of outdoor adventure activities, under controlled experimental conditions and using computer simulations.	Developed three computer-generated scenarios based on observations of 10 outdoor leaders' (25 years of experience) experiences at work. Studied how 104 leaders in outdoor education (average 14 years of experience) made decisions.	Experts relied on serial processing and intuition while less experienced leaders used comparison analysis when making decisions and familiarity with the situation to guide decision making.	Simulations cannot capture all the characteristics of field-based situations.
Chavin, Clostermann, & Hoc (2009)	To study how ship-handling training programs influenced apprentice watch officers' decisions on avoiding collisions.	Recruited 81 students, in their fifth year of training and with at approximately two years of onboard training, to use ship-handling navigator simulations.	Students knew ship-handling regulations but struggled to safely handle ships in complex ship-handling simulations. Students with greater experience were able to anticipate and select safer handling options.	Researchers did not observe or interview apprentice ship-handling watch officers' decision making in real-world situations.
Franklin, Liu, Nguyen, Johnson, Robinson, Okakor, King, Patel, & Zang (2011)	To study attending physicians' decision making during emergency department task transitions.	Used shadow observations with think aloud protocols to gather research data from five physicians.	Found that 50% of physicians' decision making was unplanned because of real-world factors they encountered.	Researchers did not assess the impact of physicians' recommendations.

Greitzer, Podmore, Robinson, & Ey (2010)	To explore the use of an NDM model to train electric power system operators how to make power-grid operation decisions.	Studied four power-grid operators' team communications to develop an NDM model. Used the NDM model to assess electric power system operators' decision making and training. Used shadow observations and a think-aloud protocol to observe the operators and to determine how they performed decision making.	The RPD model in team-based scenario training can be used to capture power system operators thought processes; measure operators' situation assessment; and identify the mental simulations used by operators.	Results are based on a preliminary test of the use of an NDM model. More research is needed.
Helsloot, & Groenendaal (2011)	To understand how Dutch forensic team leaders evaluate and utilize information.	(1) Focus group interview composed of Dutch forensic team leaders; (2) Game-based experiment: Participants were randomly separated into two groups and made decisions based on a game with three crime scenarios; (3) Control group of lay people.	During information assessments and decision making, team leaders made requests for more information; emotionally charged cases influenced decision making; relied on forensic evidence and on unverified information to make decisions.	Researchers relied on group scores, but they did not take into account group dynamics. More research is needed to validate the game used in the study.

Jenkins, Stanton, Salmon, Walker, and Rafferty (2010)	To explore whether a Cognitive Work Analysis (CWA), decision making ladder and an NDM approach can be used to design a decision support system.	Interviewed an experienced tank commander to develop atypical and typical land-based combat identification scenarios. A model was generated and validated by eight experienced tank commanders.	Used CWA to complement NDM studies. Researchers captured the relationships between information, systems, and options available for decision making.	The sample data was small (one experienced tank commander) and further research is needed on the use of CWA to complement NDM studies.
Marshall, West, Aitken (2011)	To study what information intensive care nurses used in a clinical setting and their perception of the availability and usefulness of information.	Six intensive care nurses with two to 15 years' experience participated in the study. The participants were asked to use the think-aloud protocol in their clinical practice, during the initial two hours of their shift. Follow-up retrospective interviews were also conducted.	Time constraints and information usefulness and availability determined which information intensive care nurses preferred. Nurses indicated that people provided the most useful and accessible information, while text and electronic information was less useful due to time constraints.	Researchers could not determine whether the intensive care nurses relied on information that was based on evidence to make their decisions.

Staton & Wong (2010)	To understand rational and naturalistic decision making in software design.	Used CDM to conduct 45 minute interviews with 25 software designers. Data analysis relied on content analysis and explanation-building techniques.	Software engineers relied on problem structuring, environmental cues, external goals, knowledge, experience, and a review of alternative options to guide how they made their decisions. Their decision making process was influenced by time pressures.	The study did not observe software designers' decision making in real-world situations.
Theroux (2010)	To understand women's decisions in menopausal management during their transition from peri- and post-menopausal transitions.	Interviewed seven women between 48-58 years of age experiencing peri- and post menopausal women.	Women's decisions were influenced by internal and external factors. Women utilized individual decision support when determining the risk, benefits and potential outcomes of health related decisions.	Researchers indicated that the research sample was small.
Vamjaramta and Easton (2010)	To study how managers use mental simulations to generate decisions strategies in industries with participant interaction.	Used a non-participation shadowing technique to study a business meeting recording where a sales general manager interacted primarily with a sub-ordinate manager.	Found that participants used mental simulation when generating decisions and relied on experience. Participants revealed that their decisions were influenced by assumptions about people's cultures, past encounters with people, and general business or industry knowledge.	Did not interview research participants but studied a recording of a business meeting. This did not provide the researcher an opportunity to ask probing questions for clarification pertaining to participants' decision making.

Researchers' use of interviews, observations, and simulations based on experts' field experiences found that research participants relied on their experience, knowledge, and information derived from situation assessments. The majority of the researchers found that the research participants made decisions similar to the processes described in the RPD model, such as information seeking (Helsloot & Groenendaal, 2011) and mental simulation (Baretto & Ribeiro, 2012), during initial decisions and decisions where more information or clarification was necessary (Helsloot & Groenendaal, 2011). Furthermore, researchers found that scenarios and computer-based simulations helped the researchers understand how other experts, experienced, or novice's research participants made decisions. This indicates that researchers are using data gathered using NDM techniques to understand decision making in laboratory settings. The researchers found that the research participants reported that real world characteristics influenced their decision making. Time constraints seemed to influence research participants' decision making in high-stake environments (Franklin, et al., 2011; Helsloot & Groenendaal, 2011; Marshall et al., 2011).

Researchers made recommendations for training real world decision makers, using exemplar scenarios and computer simulations to educate these decision makers on how experts perform decision making. These recommendations included increasing real-world experiences, such as onboard training for ship-handling watch officers (Chavin et al., 2009). They developed techniques for studying how real-world factors influenced participants' decision making by combining CWA's decision ladder and the RPD model. They also extended previous research, such as studies in medicine (Franklin et al, 2011; Marshall et al., 2011), military (Baber et al., 2010; Jenkins et al, 2010), industry (Greitzer, et al, 2010), business (Vamjaramta and Easton, 2010), and safety (Baretto & Ribeiro, 2012; Chavin et al, 2009). In addition, they explored

NDM research in new disciplines, such as forensic science (Helsloot, Groenendaal, 2011) and outdoor leadership (Boyes & O'Hare, 2011).

Since NDM is interested in studying the cognitive processes that shape naturalistic decision making, the NDM field has turned to research on cognition to inform its framework. NDM researchers are interested in cognitive science, situated cognition, cognitive task analysis, ill-structured problem solving, and studies on experts and novices to understand how to research the contextual and task factors that shape how we make real world decisions. Cognitive science, in general, focuses on the study of the mind. Situated cognition posits that cognition works within a given context, and cognition cannot be understood without considering the context in which the learning occurs (Brown, Collins, & Duguid, 1989; Lave, 1998; Resnick, Levine, & Teasley, 1996; Wortham, 2001). From this viewpoint, context is not passive but rather influences how learning occurs (Weldon 2001) and is used as a resource when solving a problem (Suchman, 1987).

Cognitive task analysis (CTA) leverages applied psychology; in particular, industrial psychology, to explore the guiding principles that people use to conduct their work (Hoffman, & Militello, 2009). Branch (2009) defines a task as an "apportionment of work; usually an assigned piece of work intended to be finished within a certain period of time" (p. 63). Branch also identifies three types of tasks: "Cognitive task (involves thoughts, ideas, and perceptions – Knowledge), Motor task (involves physical exertion - Skill), and Order task (involves a sequence of tasks - Procedure)" (p. 63). CTA explores the thinking processes that underlie the performance of tasks and reveals the thought processes utilized in solving problems and/or making decisions while performing a task. CTA is interested in the dynamics present in the performance of a task and in the performer's thinking processes, knowledge, and proficiencies.

CTA appears to be a useful tool in exploring naturalistic decision making. Many CTA methods exist, including the critical decision method, which my research utilized.

Many of the tasks studied in cognitive task analysis require solving ill-structured problems. A *problem* is defined as an unknown entity in some context in which its initial state differs from its desired state, and a direct, obvious way to solve a problem is not clear. In addition, there should be a need to solve the problem (Jonassen, 1997) for the unsolved question to be considered a problem.

Problems can be viewed as being either well-structured or ill-structured problems. A *well-structured problem* is defined as having "at least one problem space in which can be represented the initial problem state, the goal state, and all other states that may be reached or considered, in the course of attempting a solution to a problem" (Simon, 1973, p. 183). An *ill-structured problem* can be defined as a problem in which the initial state's requirements are ambiguous, the specifications needed to define the solution are insufficient, and there is a need to progressively define new constraints during the solution-generating process (Eastman, 1969; Jonassen, 1977; Simon, 1973). Most real world problems are ill-defined.

In the process of solving ill-structured problems, the goal is to move the problem from an unknown stated state to a known state. This is referred to as *problem structuring* (Smith, 1997). Hernandez-Serrano states that ill-structured problems "do not have a single solution, are open-ended, are composed of many sub-problems, frequently have many possible solution paths, and possess no clear beginning or end" (Hernandez-Serrano, 2001, p. 3). As noted by Orasanu and Connelly (1993), real-world decision making typically requires that the decision maker solve ill-structured problems (Zsambok & Klein, 1997).

These problems can be solved in a variety of ways. Klein, Orasanu, Calderwood, and Zsambok (1993) present a compilation of models and paradigms developed based on naturalistic decision-making theory and research as well as the challenges that NDM researchers encounter when conducting research in real-world environments. The most relevant model and paradigms to this study are the Recognition-Primed Decision model (Klein, 1993) and Orasanu and Connolly's eight characteristics of real-world decision making that have been previously discussed. Although helpful, the models are somewhat limited. They are descriptive models and, hence, cannot be used to develop procedures on how expert evaluators perform naturalistic decision making. The models, however, can shed light on how expert evaluators make decisions. In so far as the context of naturalistic decision making is critical to making the decision, the models may or may not be relevant to the particular situations that experts encounter.

Naturalistic Decision Making in the Context of Expert Evaluation

Naturalistic decision making complements expert evaluators' approach to real world decisions and addresses the gaps and challenges that evaluation researchers have identified in evaluators' decision making. Fitzpatrick, Sanders, and Worthen (2011); Patton (2008, 2010); and Stufflebeam and Shinkfield (2007), for example, explore how evaluators conduct evaluations. While these books do not focus on naturalistic decision making, they provide information about the challenges that evaluators encounter when conducting an evaluation and how evaluators should adapt to the challenges they encounter. Patton, for example, emphasizes the unique and complex characteristics of each evaluation decision as well as the importance of situation assessments in making evaluator decisions. Bamberger, Rugh, and Mabry (2006) identify budgetary, timely, and political constraints that shape evaluators' decision making. These challenges represent the key factors addressed by naturalistic decision making.

A major contribution of the NDM research to this study focuses on adaptive cognition (Anderson, 1990) and how it contributes to making dynamic decisions. Adaptive cognition refers to "the human cognitive apparatus, including perception, classification, memory, and problem solving, has evolved adaptively to cope with different kinds of tasks found in everyday environments" (Orasanu et al., 1993). Klein (2003, 2009) proposes that experts use adaptive decision making in real-world situations. In particular, Klein emphasizes naturalistic decision makers' reliance on experience to make decisions.

Tourmen (2009) and Kundin (2008, 2010) have directly studied how NDM can provide a framework to study evaluators' decision making. Tourmen analyzed how evaluators' levels of expertise affected their naturalistic decision making. Tourmen used Cognitive Task Analysis and Cognitive work analysis to explore decision making processes. Kundin also studied expert evaluators' naturalistic decision making, but Kundin used the Critical Decision Method and the Recognition-Primed Model to understand how evaluators make decisions. Tourmen and Kundin found that evaluators rely on expertise, judgment, and confidence when making decisions. Both of them found that evaluators use situation assessments to understand the contexts they encounter. It is unclear from Tourmen's and Kundin's work whether their descriptions of the evaluators' decision making context were complete. In particular, it was not clear whether all of the critical incidents that evaluators encountered were presented and whether all the decisions the evaluators made were presented. These findings helped me formulate the following questions:

(1) What critical incidents do evaluators encounter?

(2) What factors influence evaluators' decision making?

(3) How do evaluators make naturalistic decisions?

Evaluation as Naturalistic Decision Making: A Multiple-Case Study

Fitzpatrick's, Christie's, and Mark's (2009) goal in writing *Evaluation in Action: Interviews with Expert Evaluators* was to help "readers learn more about the choices an evaluator makes in the course of planning an evaluation" (p. 542). An analysis of the interviews reveals that the expert evaluators considered the NDM key contextual and task factors in making their decisions. The interviews revealed (a) evaluation as an ill-structured problem; (b) evaluation's complex and dynamic nature and the importance of context in evaluation; (c) competing/shifting stakeholder goals; (d) action/feedback loops in evaluation; (e) the presence of time stress in evaluation; (f) high stakes in evaluation; and (g) the evaluator's need to navigate through organizational goals in evaluation practice, all of which are aspects present in naturalistic decision making.

Evaluation as an ill-structured problem. Christie and Rose (2003) stated in the article, "Learning about Evaluation through Dialogue: Lessons from an Informal Discussion Group" that "practices are highly constrained by contextual and structural factors" (p. 91). The questions evaluators are trying to answer, as well as the problems they confront, are often complex. In real-world situations, many of these problems are ill-structured (Zsambok & Klein, 1997).

As noted by Rossi, Lipsey, and Freeman (2004), evaluators often facilitate the identification of a program's theory of change or the program theory. They also sometimes intentionally or unintentionally assist with the design, refinement, or implementation of the evaluand as they determine evaluation questions, work with stakeholders to generate an evaluation plan, design data-gathering instruments, and determine the methodology used to conduct the evaluation. For example, when seeking ways of collecting evaluation data in the context of e-learning programs, evaluators may reveal a need to modify the evaluand in such a

way that allows certain types of information to be collected by a built-in tracking program (Reeves & Hedberg, 2003).

In addition, evaluators play many roles depending on the context of the evaluation. For example, in Fetterman's program evaluation, his focus was on the "processes and discrepancies" (p. 366) of the Stanford Teacher Education Program's model or theory of change. While helping the Rockefeller Foundation formulate program evaluation questions, "Wallis and Dukay's team focused on outcomes" (p. 366) as they evaluated a children's orphan center in Tanzania. The many roles that evaluators perform when conducting an evaluation is best summed up by Bloodsoe's statement,

> I believe the role of the evaluator changes depending on the situation. I think the role is multifaceted. Sometimes you are the judge and jury. Other times you are a facilitator. Still other times, you're an advocate. And still others, you may find yourself in program development. It depends on the context, the resources, and the desires of the stakeholders. To limit evaluators to one facet of the position is to disregard the complexity of it. (p. 316)

Evaluators often deal with ethical issues that challenge both their own norms and the norms of the organization. Evaluators should work with multiple stakeholders, each with their own agendas. All these aspects were revealed in the interviews conducted by Fitzpatrick, Christi, and Mark (2009). Each of the evaluators had a different understanding as to how an evaluation should be conducted.

Dynamics and complexity in evaluations. Evaluations can be complex, as was revealed in an interview with Stewart Donaldson, who evaluated a Work and Health Initiative funded by the California Wellness Foundation. Donaldson stated that "some programs are more complex than others. Most human programs in this [Work and Health] Initiative were complex and multifaceted. The effects are subtle. They are diffuse" (Fitzpatrick, Christie, & Mark, 2009, p. 224). The dynamics in an evaluation context can change; in particular, group dynamics can

change. Donaldson also stated that there can be a "good relationship with clearly defined roles," there may be "dissention among the [evaluation group]," or the evaluand stakeholders may not "like the evaluation team" or "like being evaluated" (p. 224).

Context in evaluation. Stake (1995) stated that evaluators should be able to change their approaches as necessary to adapt to the program context and stakeholders' needs. All the interviews reviewed in Fitzpatrick et al. (2009) highlighted that evaluators utilize contexts to inform their decision making process. Fitzpatrick (2009) stated in the chapter "Analysis, Interpretations and Conclusions" that "Each of these evaluators' choices . . . was influenced by the context of their work – the stage of the program, its characteristics, the expectations and values of the stakeholders, and the nature of the surrounding environment" (p. 363).

Competing and shifting stakeholder goals. The interview with Riccio indicated that evaluation stakeholders could have competing goals. This is highlighted in Riccio's statement regarding the legislators who commissioned the evaluation of the Welfare-to-Work program that "Liberal and conservative legislators were advocating reform bills that pushed in different directions" (p. 30). Bickman's interview further highlighted competing and shifting goals. As Bickman stated,

> That is what the Department of the Army wanted in the beginning — just to look at the cost to them. They wanted the right to approve anything I published, which I refused. Then they wanted to be able to comment on anything we published, which I explained was not under their control . . . . It was a battle with the Army throughout the project to maintain the integrity of the design. (p. 73)

The Army's and the evaluator's goals became more defined as the issues regarding "control" were discussed.

Multiple Stakeholders and Action/Feedback loops in evaluations. Fitzpatrick's (Fitzpatrick et al., 2009) interview with Wallis and Dukay highlighted the importance of

feedback from multiple stakeholders when conducting an evaluation. As Dukay stated, "Without such understandings of the cultural context we could have easily misinterpreted what we were hearing in the interviews and focus groups. So having psychiatrists, anthropologists, and sociologists from Tanzania was essential to the conduct of the study" (p. 332).

The use of previous evaluations can also be considered an aspect of the feedback loop. As highlighted in Fitzpatrick et al. (2009), Ross Connor's evaluation of the Colorado Healthy Communities Initiative received "recognition for developing community indicators that could be used by other programs to help improve the health of communities" (p. 382). Fitzpatrick's reflections on an interview with Len Bickman, who conducted an evaluation of Fort Bragg and Stark County's System of Care for Children and Adolescents, highlights the importance of learning from previous evaluations. Fitzpatrick stated that "one of the elements [she] admired most [about Bickman] was his conscious effort to build and learn from previous evaluations" (p. 90).

Time stress in evaluations. The lack of sufficient time to conduct evaluations is present in almost all the interviews. Stress due to insufficient time is highlighted when James Riccio shared what he experienced during the evaluation of California's Welfare-to-Work program. Riccio noted that evaluators should often work under significant time constraints. For example, Riccio stated, "It [was] difficult to anticipate the amount of time and energy that must be spent explaining and building local support for random assignment" (p. 40).

Fitzpatrick et al. (2009) also interviewed Jennifer Greene, who conducted an evaluation of a natural resources leadership program. Greene also highlighted the importance of timing, especially in reporting findings. Greene stated, "By the time this report [finally] came out, the

moment for meaningful engagement with these issues had passed, which is one reason not to wait to do your important evaluation reporting until the end" (p. 62).

High stakes in evaluations. The high stake aspect of program evaluation is highlighted in the authors' interview with David Fetterman, who conducted an evaluation of Stanford's Teacher Education Program (STEP). Fetterman asserted that the evaluation had a high impact on stakeholders. He stated, "I was interviewing one colleague when suddenly, in the middle of the interview, the guy was almost in tears about being pushed away from teaching in the [STEP] program" (p. 103). As shown in the interview with Allan Wallis and Victor Dukay, evaluators of a children's orphan center in Tanzania, the evaluation determined that orphans at the center were "healthier, liked school more, [and] had higher educational aspirations than the village orphans or children with their parents" (p. 331). These were high stakes findings that impacted the welfare of the center's children.

Navigating through organizational norms and goals. Fitzpatrick's reflections on Katrina Bledsoe's interview highlighted an evaluator's need to navigate through an organization's goals and norms. As Fitzpatrick stated, "Bledsoe says that she achieved [evaluation] use by framing her recommendations in terms of what was feasible to the organization and suggesting the organization was moving in the direction of the recommendations already" (p. 386). Since program evaluations are conducted within profit and non-profit organizations, evaluators must work within an organization's norms and goals. The authors of the book highlight "that the evaluator should model the behavior they would like to see in stakeholders by receiving their suggestions without defensiveness and making suggestions as appropriate" (p. 389).

Conceptual Frameworks Used in This Study

The NDM naturally lends itself to the goal of this study, which is to study expert evaluators' decision making processes in the context of HIV/AIDS health education programs. The Recognition-Primed Decision Model (Klein, 1997) and the Conceptual Framework for How Evaluators Make Practice Decisions (Kundin, 2008, 2010) represent two key models that informed this study and guided me in generating the interview questions, determined how the data should be analyzed, and provided the lens through which to interpret how evaluators make naturalistic decisions. The next two sections provide an overview of the Recognition-Primed Decision Model and the Conceptual Framework for How Evaluators Make Practice Decisions (Kundin, 2008, 2010). This overview is followed by the framework that guided my research.

The Recognition-Primed Decision Model

Klein (1997) has modeled naturalistic decision making in the Recognition-Primed Decision Model. This model is contrasted with the rational decision making approach in which decision makers gather all the available information; conduct exhaustive, concurrent analyses of the available options; and then chooses the optimum solution (Hutton & Klein, 1999). This model focuses on expert decision making. In doing so, it recognizes the unique manner in which experts perform decision making. Novices often need to go through iterations that compare solutions, often use a time-consuming trial-and-error process, or opt for an insufficiently informed first choice, whereas expert/experienced decision makers quickly devise plans and rapidly determine whether they will work. Klein (Zsambok & Klein, 1997) found that experts' vast experience does not slow them down; instead, it allows them to make quick decisions by recognizing patterns.

The Recognition-Primed Decision Model (RPD, see Figure 2.1) depicts Klein's research (Klein, 1997). The RPD model is a naturalistic decision making model that describes how experienced people or experts make decisions in real-world settings. The RPD model shows that experts or experienced decision makers utilize situation analysis and mental simulation to determine whether a situation is typical and to evaluate their decisions. The RPD model proposes that people rely on situation assessments and mental simulations to formulate possible courses of action when they encounter complex and dynamic situations. It also proposes that situation assessment and mental simulations are used to evaluate the feasibility of a course of action.

Decision makers' previous experiences provide them with the skills necessary to identify patterns. According to Klein, these patterns are informed by Recognition's Four-by Products: goals, expectations, cues, and typical actions (Klein, 1997). Klein's research also found that decision makers generate several alternative solutions, but that they sequentially evaluate "each choice on its own merits even if they cycle through several possibilities. [Klein also found that decision makers] … don't need the best solution. They just need one that works" (Klein as quoted in Breen, 2000, p. 3 para. 10).

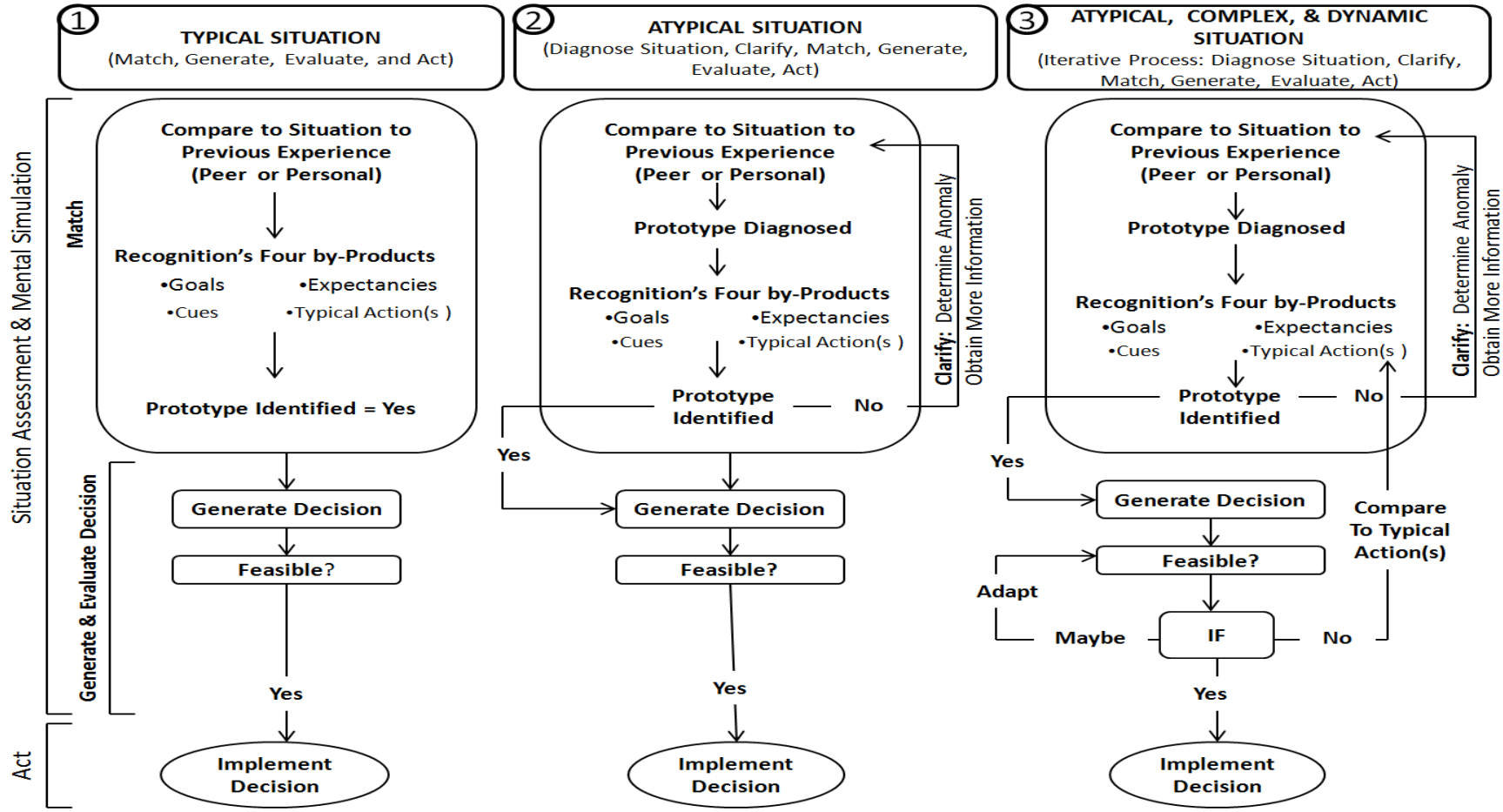

Figure 2.1. Recognition-Primed Decision Making Model. Adapted from "Recognition-Primed Decision Model (RPD): Looking Back, Looking Forward" in C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285-292), New York, NY: Routledge. Copyright 1997 by Taylor and Francis. Adapted with permission.

In the RPD model, three levels are depicted. Level one depicts decision making where a situation is familiar, and typical actions can be implemented [prototype]. Level two depicts decision making when an atypical situation is encountered, and pattern matching is an iterative process. In level three, the decision maker encounters a situation that is atypical, complex, and dynamic. In level three, there is a need to conduct multiple situation assessments, generate multiple decisions, and conduct multiple evaluations of proposed decisions. In my study, I expect to encounter the three levels depicted in the RPD model.

The RPD model has its limitations. The RPD model is used to understand experienced or expert decision making and is not usually applied to studying novice's decision making. The model does not address issues related to decision makers' memory or metacognitive processes. Until recently, the RPD model did not address how decision making is influenced by team dynamics and organizational norms (Zsambok & Klein, 1997).

Conceptual Framework for How Evaluators Make Practice Decisions

Kundin's (2010) Conceptual Framework for How Evaluators Make Practice Decisions (see Figure 2.2) was utilized in this research to understand the critical incidents HIV/AIDS education program evaluators encounter, the factors that influence their decision making, and the way they make decisions. Kundin's (2008) model was based on an "NDM study of the decisions of evaluators who have at least five years of evaluation experience" (p. 48). Kundin's (2008, 2010) framework is based on the following: (1) the definition of evaluation context (Greene, 2005); (2) theories that evaluators use based on evaluation practice; (3) research on working logic in use frameworks (Fournier, 1995); (4) logic in action (Hansen, 2005); (5) research on reflective practice (Schön, 1983); and (6) research on evaluation practice by Schwant (2005).

Kundin's (2008, 2010) model relies heavily on Endsley's (Ensley & Garland, 2000) research conducted on situation awareness.

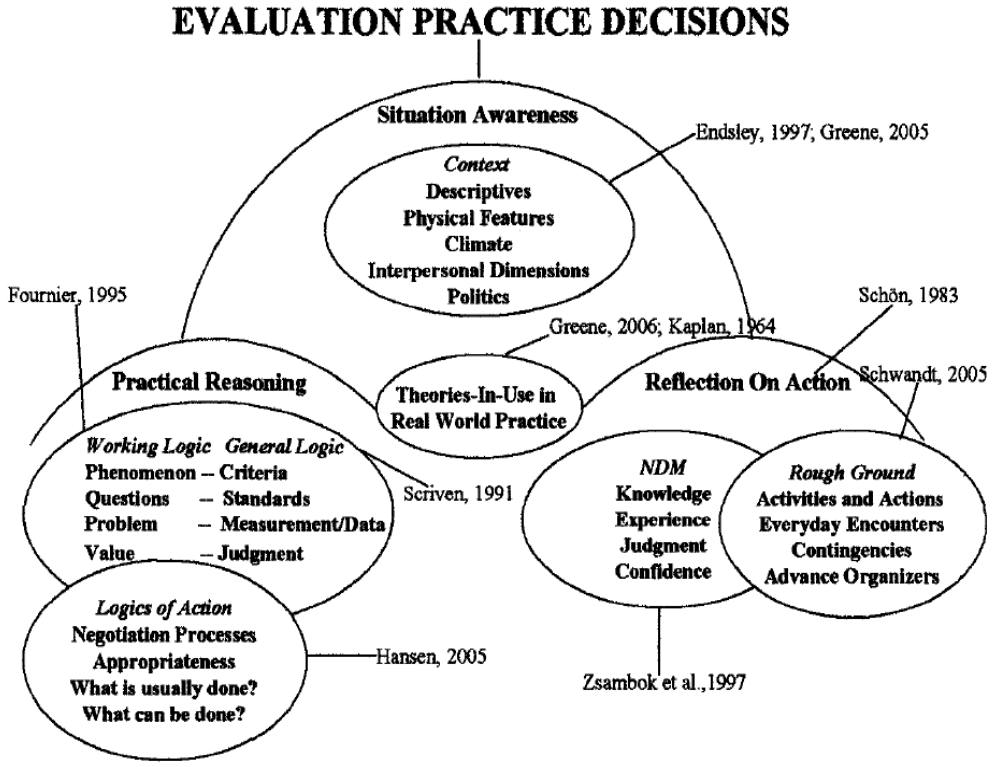

Figure 2.2. Conceptual Framework for How Evaluators Make Practice Decisions. From "A Conceptual Framework for How Evaluators Make Everyday Decisions" by D. M. Kundin, 2010, *American Journal of Evaluation, 31*(3). p. 349. Copyright 2010 by Sage. Reprinted with permission.

Kundin notes that an evaluator's ability to conduct a situation assessment affects the evaluator's practice decisions. In addition, evaluators use previous "knowledge, experience and judgments" (p. 351) in the selection of the models they utilize or develop to conduct evaluations. Kundin (2010) cites Greene's (2005) research in the context where evaluation practice occurs as

> (a) the descriptive and demographic character of the setting, (b) the material and economic features of a setting (buildings, gathering spaces, resources and

> technology, (c) the institutional and organization climate in a setting (agency, public institution and private business), (d) the interpersonal dimensions of a setting (norms that frame and guide relationships), (e) the political dynamics of a setting. (p. 350)

The second element in the model refers to practical reasoning. This element identifies the working logic used to generate possible solutions after conducting a situation analysis. In Kundin's model, "general logic refers to the logic used to choose which activities to address when conducting an evaluation" (p. 353). Working logic is used to "describe the diverse considerations that practitioners give to a particular phenomenon (evaluation situation) . . . and an argument structure that reveals a pattern of reasoning that supports one's conclusions" (p. 353).

In level two, Kundin highlights evaluators' use of "logic of action" "to determine the 'appropriateness' of their evaluation approach based on their situation assessments" and to determine what can be negotiated where compromises are needed" (p. 352). When faced with typical situations evaluators ask themselves, "What is usually done?" (p. 352). When they encounter atypical situations evaluators ask themselves, "What can be done?" (p. 352). The question "What is usually done?" is closely aligned with what Kundin (2008, 2010) refers to as "Theories-in-Use in Real-World Practice" or models "based on profession judgments and are used or proposed by either the practitioner or other evaluators . . . but can only be realized in a specific content with real demands and limitations" (p. 353).

Schön's (1983) work is referred to in the model section titled *reflection-in-action*. Reflection involves re-thinking and challenging assumptions that are used to address a new problem that is being encountered. Usually, a reflection occurs when

> there is some puzzling, or troubling, or interesting phenomenon with which the individual is trying to deal. As he tries to make sense of it, he also reflects on the

> understandings to which have been implicit in his action, understandings which he surfaces, criticizes, restructures, and embodies in further action. (p. 50)

It is often referred to as "'thinking on your feet,' 'keeping your wits about you,' and 'learning by doing'" (p. 25).

According to Ghere, King, Stevahn, and Minnema (2006), reflective practice is one of the essential competencies for program evaluators. Reflective program evaluators should "focus on understanding one's practice and level of evaluation expertise, including an awareness of the need for professional growth" (p. 110). This section on reflection-in-action in Kundin's model also draws on naturalistic decision making research by Zsambok and Klein (1997), which is used to depict evaluators' use of knowledge, experience, judgment, and confidence. It also highlights the evaluators' ability to work with tools in dynamic and complex evaluation environments, based on Schwant's (2005) position that evaluators must often conduct evaluations in "rough ground" when decisions are made in spaces in which there exist "distinctive tensions, contradictions, paradoxes and dilemmas that affect our understanding of self, world and other and consequently our practices" (p. 361). The evaluators also utilize their Theories-in-Use in Real World Practice and their Practical Reasoning to assist them in decision making.

Kundin's (2008) research focuses on "exploring the reasoning evaluator's use when making decisions as they go about their everyday work." (p. 2). Kundin (2008) interviewed eleven evaluators and interview quotes to support the research findings. Kundin's study fails to include the cases or narratives from the research interviews, which would have added to "the deep rich" description that narratives from participants' interviews could bring to a study. The context is considered as having a major influence on how evaluators conduct evaluations (Patton, 2008). The use of narrative research would have assisted in understanding the context in which

evaluators in Kundin's study made their decisions when conducting evaluations and would have provided a further insight into the dynamic decision making process.

Guiding Framework for this Research

The guiding framework for my research incorporates the NDM framework and the two models depicted in Figures 2.3 and 2.4. In Figure 2.3, the different layers in the planning phase of an evaluation in real-world settings are depicted. I decided to focus on the evaluation planning phase because the end product of the planning phase guides the other phases in an evaluation's life cycle (Fitzpatrick, et al., 2010). Kusek and Rist (2004) recommended that evaluations should be conducted in programs in various life cycles. I added this aspect to the model, because the implementation of an evaluation in a program's life cycle can influence an evaluator's approach to the evaluation (Stufflebeam & Shinkfield, 2007). Evaluators attempt to implement evaluations that are systematic and as objective as possible (OECD, 1999). I anticipated that evaluators would find conducting a systematic and objective evaluation challenging, so I added it to the guiding framework.

Evaluators use criteria to judge the merit or worth of a program. The criteria are as follows: efficiency, effectiveness, impact, sustainability, fulfillment of objectives, and relevance (OECD, 1999). I decided that it was important to add the evaluation criteria to the guiding framework, as the criteria will likely have an influence on how the evaluation is implemented (Fitzpatrick, et al., 2010).

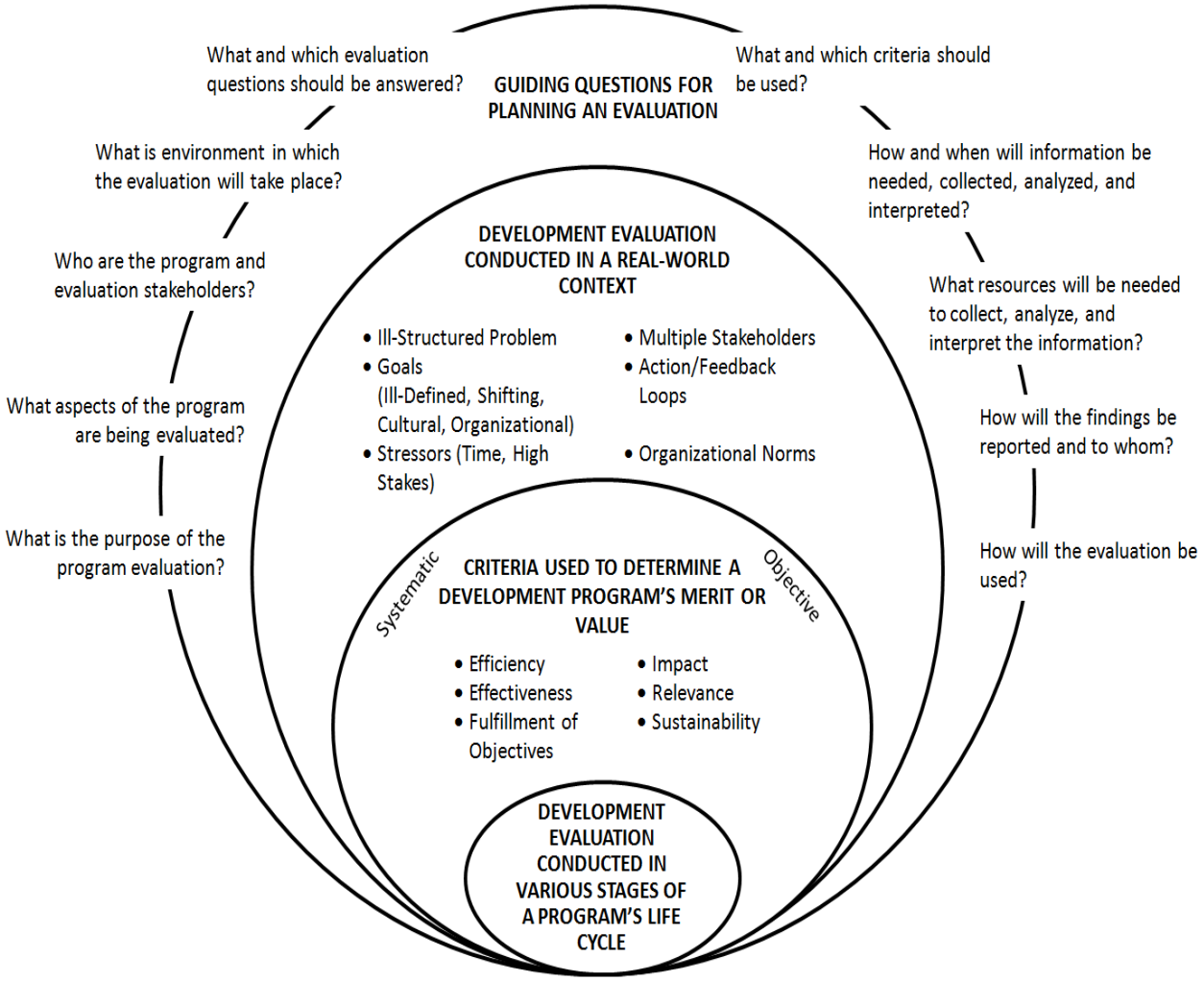

Figure 2.3. Planning an evaluation of a development program.

Since I am studying how evaluators conduct evaluations in the field, I decided that it was necessary to include the characteristics found in real-world settings (Zsambok & Klein, 1997). The real-world decision making components are categorized as (a) goals, (b) ill-structured problems, (c) stressors, (d) multiple players/stakeholders, (e) organizational norms, and (f) action/feedback loops. Further depicted are the typical questions evaluators may ask when planning an evaluation (Fitzpatrick et al., 2010). These guiding questions were included, as these questions influence how evaluators plan an evaluation.

The second section of the guiding framework (see Figure 2.4) is informed by the theories found in the Recognition-Primed Decision Making (RPD) Model (Klein, 1997) and Kundin's (2008, 2010) Conceptual Framework for How Evaluators Make Practice Decisions' (CFHEMPD). The RPD model was used as a guiding framework in Kundin's research (2008, 2010). Kundin's research revealed that experienced evaluators conduct situation assessments to help them understand how to solve typical or atypical situations where a problem must be solved. Furthermore, Kundin's research found that the evaluators' situation assessments are informed by their personal experiences or peer-based experiences. Kundin's work also found that evaluators also rely on their knowledge, judgment, and confidence when making practice decisions.

After studying the RPD model and the CFHEMPD, I determined that evaluators would likely utilize situation assessment and mental simulation to assist them in their decision making. I anticipate that an evaluator's naturalistic decision making is likely to occur in three major phases. The first phase will focus on conducting a situation assessment to determine whether he or she has encountered a similar situation [prototype].

During a situation assessment, the evaluator will attempt to compare the critical incident to a prototype. The matching will be informed by Recognition's Four by-Products (Klein, 1997) which are goals, expectations, clues, and typical actions. The evaluator will continue the match and compare process until he or she is satisfied with a selected prototype. Then, the evaluator will use a selected prototype to sequentially generate possible decisions.

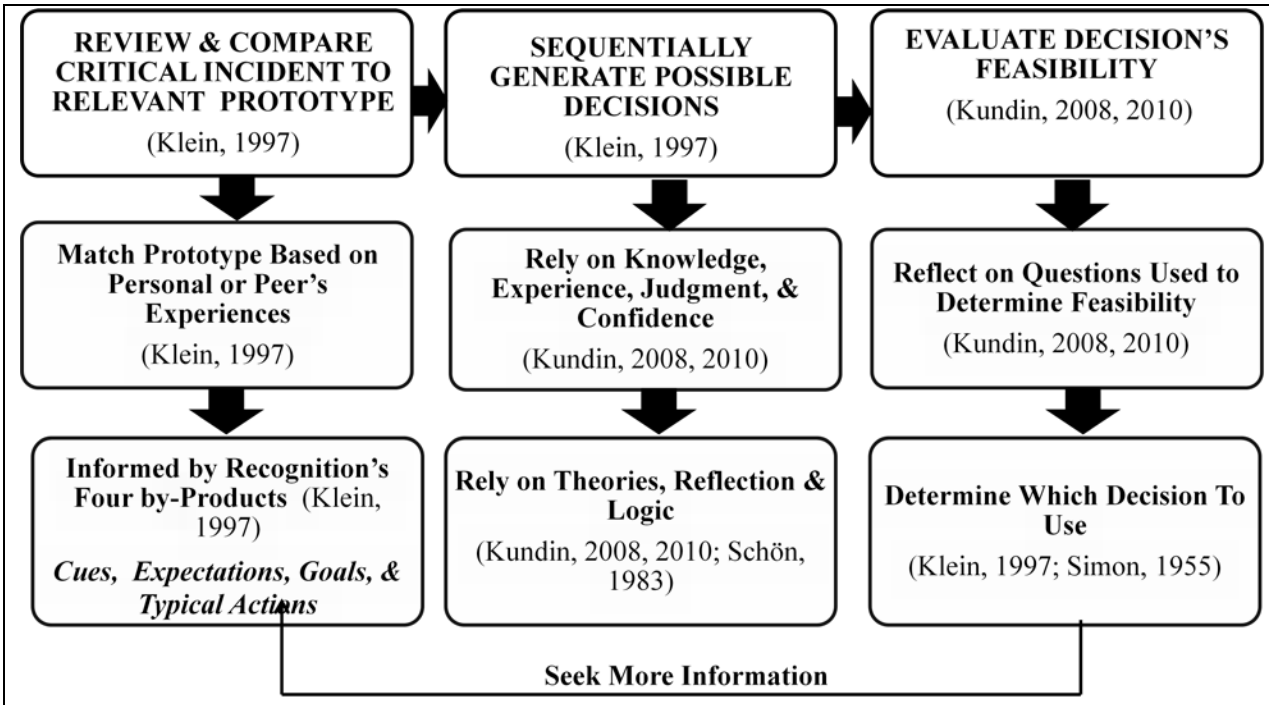

Figure 2.4. Conceptual framework of program evaluators' naturalistic decision making process during critical incidents.

In the next phase, the generated decisions are evaluated. The evaluators' assessment of whether the decision will solve the problem is guided by their knowledge, experience, judgment, and confidence (Klein, 1997; Kundin, 2008, 2010). The evaluators rely on reflection and logic (Kundin, 2008, 2010; Schön, 1983) to determine whether their decisions will help solve the problem. In the final phase, evaluators assess the feasibility of implementing their decisions. Kundin's (2008, 2010) research indicates that evaluators reflect on questions such as, "Is this an appropriate decision?" Evaluators are expected to use reflection (Schön, 1983) and an if-then-else cyclical process (Klein, 1997) to determine which decision is to be used.

Chapter Summary

This chapter provides an overview of the evaluation and decision making theories and literature that guided my research. Since the focus of my study is real world decision making by

real world experts, naturalistic decision making provided a sound framework for my research. In addition, it provides a multiple case study that illustrates evaluators' naturalistic decision making. Kundin's (2008, 2010) Conceptual Framework for How Evaluators Make Practice Decisions and Klein's (1978) Recognition-Primed Decision Model created a foundation for developing the guiding framework of my study. A description of the guiding framework for this study is provided. The methodology used to identify the factors is described in Chapter Three.

CHAPTER 3

METHODOLOGY

The purpose of this qualitative exploratory study was to understand how evaluators make naturalistic decisions. Five factors contributed to creating the conditions conducive to this study:

- Research on evaluators' naturalistic decision making has begun (Kundin, 2008, 2010);
- There is limited knowledge regarding the types of critical incidents that evaluators encounter (Kundin, 2008, 2010);
- A knowledge gap exists among the factors that influence evaluators' naturalistic decision making;
- There is a continued call for further research on evaluation practice (Christi, 2009); and
- Evaluators are encouraged to become reflective practitioners (Stevahn, King, Ghere, & Minnema, 2005).

Research Design

The research design I used in my study consisted of a collective case study of seven participants' retrospective, semi-structured interviews. The unit of analysis was the individual research participant's recalled experience of an evaluation of an HIV/AIDS health education program. I used the Critical Decision Method (CDM), Phenomenography, and Narrative Analysis to study the evaluators' naturalistic decision making. The seven co-constructed member checked decision making timelines and narratives were generated from analyzed field

notes based on each research participant's two one-hour interviews. I reviewed rough drafts of the decision making timelines and narratives to generate more context-specific interview questions and elicit further clarification of the participants' experience during the second interview. This study followed Creswell's (2007) recommendations on how to conduct a narrative analysis. As recommended, an inductive analysis approach was used to analyze the decision making timelines and narratives. Phenomenographic methods were also used to generate and inform models of expert evaluators' decision making.

As stated by Stake (1995), three key differences exist between qualitative and quantitative research:

> (1) the distinction between explanation and understanding as the purpose of inquiry, (2) the distinction between a personal and impersonal role for the researcher, and (3) a distinction between knowledge discovered and knowledge constructed. (p. 37)

A qualitative research approach was utilized to answer the research questions (See Chapter 1). In qualitative research, "the researcher's intent, then, is to make sense (or interpret) the meanings others have about the world" (Creswell, 2007, p. 21). This study followed Creswell's description of the "five philosophical assumptions [that] lead to an individual's choice of qualitative research: ontology, epistemology, axiology, rhetorical, and methodological assumptions" (2007, p. 15). In Table 3.1, a list of Creswell's description of qualitative research's philosophical assumptions and how they were followed in this study are provided.

Generally, this type of research is conducted within its milieu. It should be noted that my research was not conducted at the same time that the evaluators were conducting evaluations. Consequently, I was not able to observe how evaluators "behave and act within their context" (Creswell, 2007, p. 37). Instead, retrospective interviews provided me an opportunity to study how evaluators made decisions when "in the field."

Table 3.1

Qualitative Philosophical Assumptions and How They Were Addressed

Qualitative Philosophical Assumption	Creswell's Description	How This Study Addressed the Qualitative Philosophical Assumption
Ontological	There are multiple realities that are subjectively defined through the research participants' perspectives.	The narratives revealed the research participants' perception of how real-world or naturalistic decision making occurs in the field.
Epistemological	There is a collaboration between the researcher and research participant during research.	A collaboration occurred during the co-construction of decision making timeline and narratives.
Axiological	Bias is present when conducting research, and the research environment is replete with value-driven concepts.	An audit trail was used to reveal the research participants' and the researcher's values. The researcher's bias is revealed through an audit trail, and the research participants' bias is revealed through the narratives.
Methodological	Research findings are unfolding.	A qualitative exploratory approach was utilized to reveal complex variables that are not easily delineated (Merriam, 1998), and where "the unit of analysis . . . comes into focus as the research progresses" (VanWynsberghe & Khan, 2007, p. 90). The researcher utilized an inductive approach to analyze the data.
Rhetorical	Utilization of qualitative terminology and literary devices	Literary devices were used in the narratives generated from the interviews and cases. Qualitative terminology was used to describe the research study and findings.

Note: Adapted from *Qualitative Inquiry and Research and Design: Choosing Among Five Approaches* (p. 17), by J. W. Creswell, 2007, Thousand Oaks, CA: Sage. Copyright 2007 by Sage. Adapted with permission.

Research Approach

This qualitative research utilized the critical decision method in combination with case studies, narrative analyses, and phenomenographic methods. Naturalistic research provides the

opportunity to study a subject's experience in the phenomenon's context without requiring the researcher to control for context variables (Zsambok & Klein, 1997). Previous research has shown that context and previous knowledge have a pivotal influence on how evaluators conduct evaluations (Shadish, Cook, Leviton, 1991) and, consequently, how they make decisions.

Critical Decision Method

This researcher studied expert evaluators' decision making within the framework of naturalistic decision making research (NDM). Cognitive task analysis (CTA) is one method of conducting NDM research. CTA tries "to capture what people are thinking about, what they are paying attention to, the strategies they are using to make decisions to detect problems, what they are trying to accomplish, and what they know about the way a process works" (Crandall et al., 2007, p. 10). CTA can study interviewees' stories, which makes it appropriate for qualitative research. The three aspects of CTA are "knowledge elicitation, data analysis, and knowledge representation" (p. 10).

I decided to use the Critical Decision Method (CDM) to study evaluators' naturalistic decision making, because CDM is "designed to study decision making in naturalistic settings by researching actual incidents" (Crandall et al., 2007, p. 84). Many researchers are using CDM to carry out incident-based retrospective interviews to elicit experts' stories and decisions centered on a specific challenging incident. In Figure 3.1, a high-level depiction of Cognitive Task Analysis's three phases (knowledge elicitation, knowledge representation, and data analysis) and the utilization of the critical decision method in my research are provided.

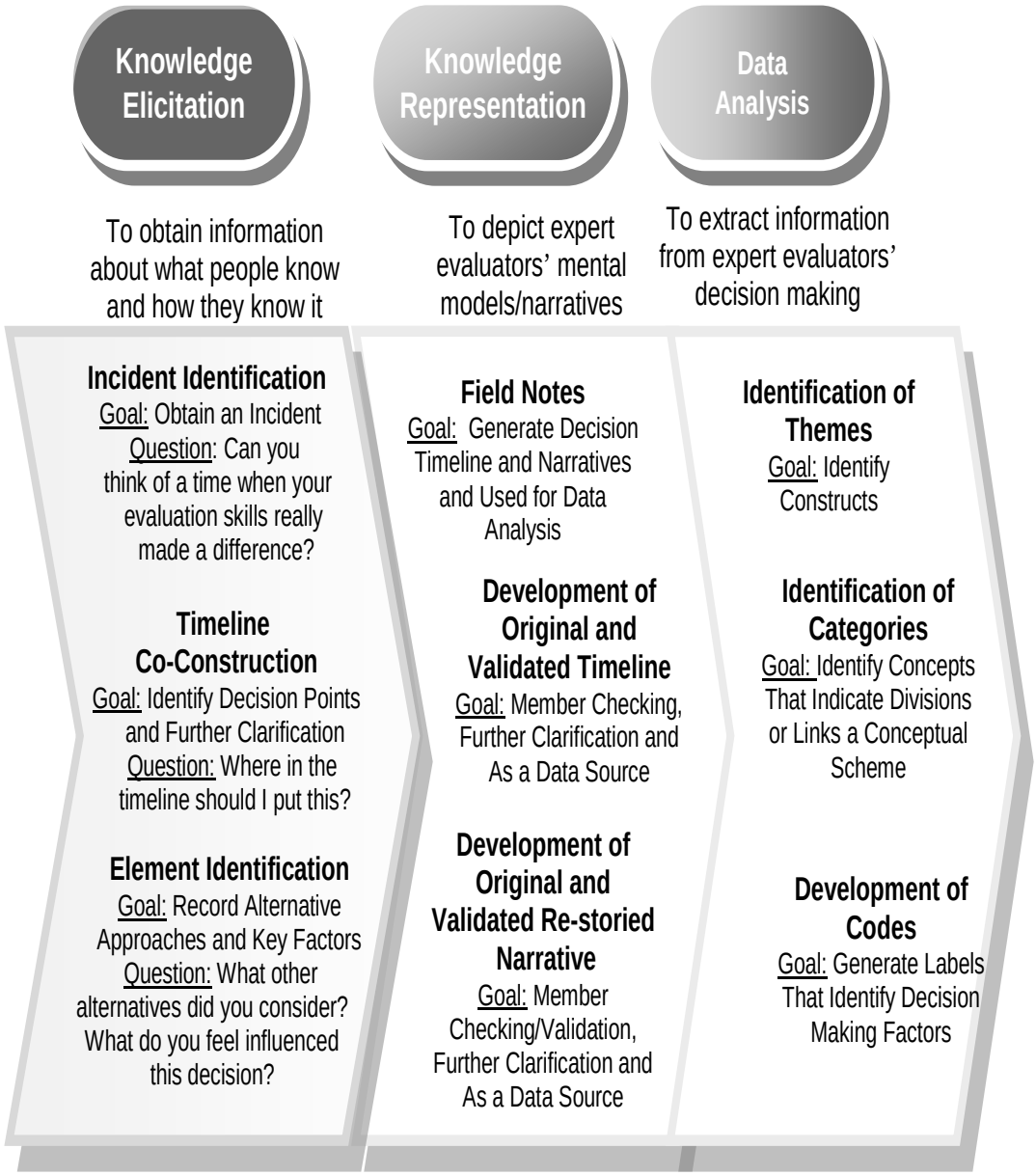

Figure 3.1. Cognitive Task Analysis' three phases and the Critical Decision Method. Adapted from *Working Minds: A Practitioner's Guide to Cognitive Task Analysis* (p. 181), by B. Crandall, G. Klein, and R. R. Hoffman, 2006, Cambridge, MA: MIT Press. Copyright by Massachusetts Institute of Technology. Adapted with permission.

Collective Case Study

Case study is a methodology that "investigates a contemporary phenomenon within a real-life context, especially where the boundaries between phenomenon and context are not clearly evident" (Yin, 2003, p. 13), such as studies that involve naturalistic decision making. Yin also stated that case studies provide the opportunity to describe a "complex social phenomenon" (p.3), such as evaluators evaluating HIV/AIDS health education programs. Instrumental cases are cases chosen to investigate the issue being researched (Stake, 1995) or, in this instance, the unit of analysis, an evaluator's decision making in the researched context.

Collective case studies do not focus on one case, but on many cases. As stated by Stake (1995), it may be beneficial to study unusual situations, because they can reveal nuances often missed in routine situations. Similarly, the critical decision method focuses on atypical events that occur when performing a task (Zsambok & Klein, 1997). In addition, collective case studies provide the opportunity to identify similarities and differences between different instrumental cases. The goal of this study was not to provide "a statistical basis for generalizing," but the goal was to provide a "refinement of understanding" of the cases being studied (Stake, 1995, pp. 6-7).

Narrative Inquiry

Narrative studies recognize that "all people have stories to tell" (Creswell, 2007, p. 119). Schank (1990) highlights that "all we know is embodied in stories" (p. 189). In addition, Schank describes an expert as a person who has a specialized knowledge based on "stories indexed well enough to find the right one at the right time" (p. 109). In this study, the individuals were experienced evaluators of HIV/AIDS health education programs with stories to tell about the critical incidents they encountered.

The narratives generated from this research were reorganized to chronologically unfold with a beginning, middle, and end. Creswell (2007) referred to this process as restorying (p. 57). The narratives were either first-order or second-order narratives in which "individuals tell stories about themselves" (Elliot, 2005, p. 12). Semi-structured interview questions were utilized to allow research participants to share stories about themselves, the critical incidents they encountered, and their decision making processes (Crandall, Klein, Hoffman, 2007), therein revealing the key factors that influenced their decision making.

Narrative analysis recommends that experiences should be studied from a dualist rather than a subjective point of view Creswell (2007). This is in contrast to phenomenography, which recommends studying experiences based on an event from a subjective point of view. I decided to follow Creswell's recommendation to use a dualist point of view to co-construct and member check the timelines and narratives generated in this study. Phenomenographic methods were used to understand the variance between the critical incidents and the factors that influenced the evaluators' decision making. Collective case study techniques provided an opportunity to study the uniqueness of each evaluator's experiences and to compare them with each other.

Phenomenography

Similar to phenomenological research (Merriam, 1998), phenomenographic methods study experiences based on an event. Phenomenography is also interested in studying decision making from the decision maker's point of view (Bogdan & Biklen, 1992). Since this study focused on evaluators' decision making, these interests were aligned.

A phenomenographic researchers' interpretivist paradigm embraces the ontological assumptions that people experience, perceive, and are aware of the world in different ways. The phenomenographic researcher's focus is not on the phenomenon but on the interplay between the

actors and the phenomenon (Bowden 2005; Marton & Booth, 1997). For example, in a phenomenographic interview, the researcher is not studying his or her own situational awareness. Investigators who use phenomenographic methods put aside their preconceived opinions about the phenomenon (Merriam, 1998). My research was exploratory, and preconceived notions were "bracketed" (Ashworth & Lucas, 2000) as much as possible. Bracketing was difficult (Mitchell et al., 1997), because previous research had informed this study. For example, my research builds on The Conceptual Framework for How Evaluators Make Practice Decisions (Kundin, 2008, 2010) and on the Recognition-Primed Decision Model, as previously described. Bracketing was also difficult, because during the co-construction of the narratives, I needed to "analyze or pre-categorize the data during interviews" (Ireland, Tambyah, Zui, & Harding, 2008, p. 7).

Research Design Procedure

This study's research design and the graphical representation of the CDM process can be observed in Figure 3.2. Preliminary activities that took place before the study began included my obtaining IRB approval before recruiting research participants. Data were gathered by using an interview protocol and interview questions. Both the interviewees and I signed IRB approved consent forms (see Appendix B).

Sweep One and Two. This study started with Sweep 1: the introduction of the study and an explanation of the consent forms. In Sweep 2, the data collection began with identifying a critical incident that revealed an expert evaluator's decision making. The identified incident could contain multiple critical incidents. The incidents should have had an impact on the results of the program evaluation. The incident must should have had the potential to reveal the

elements of expertise as well as its related cognitive aspects. The research participant's account became the foundation for the rest of the follow-up interview questions.

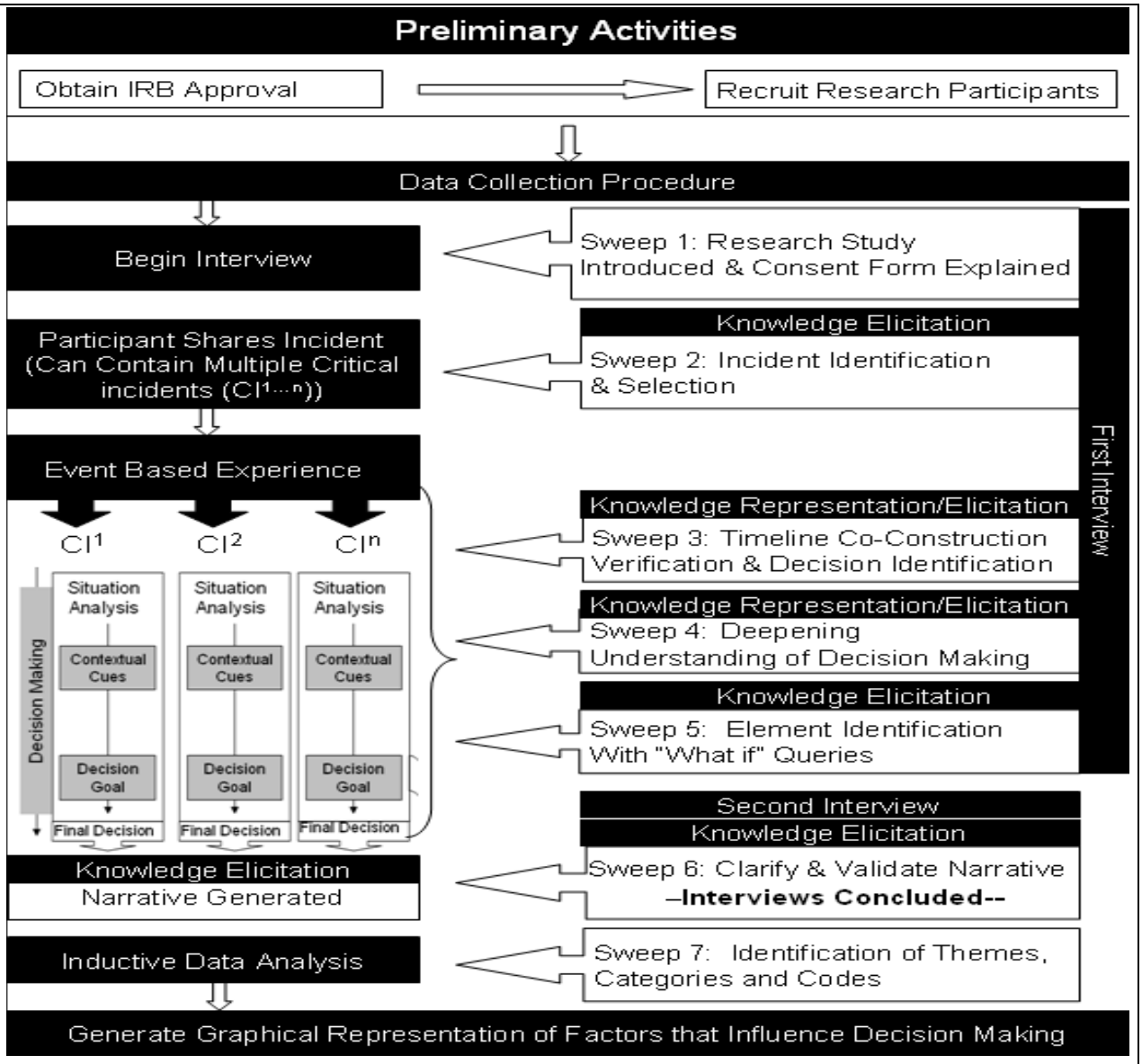

Figure 3.2. Research design and procedure. Adapted from *Working Minds: A Practitioner's Guide to Cognitive Task Analysis* (p. 74), by B. Crandall, G. Klein, and R. R. Hoffman, 2006, Cambridge, MA: MIT Press. Copyright by Massachusetts Institute of Technology. Adapted with permission.

Sweep Three. In Sweep 3, the research participant and I generated a timeline. The timeline was based on the research participant's account, his or her situation analysis of the incidents, a description of the evaluation context, the clues used to make decisions, decision goals, and final decisions. With the assistance of interview prompts, the research participant provided additional critical incident information by indicating the sequence of events and identifying decision points. The co-constructed decision making timeline assisted in examining the evaluator's actions, perceptions, thoughts, and decisions. Examples are provided in the pilot study (Appendix A), Chapter 5, and Appendix D.

The decision making timelines generated in Sweep 3 are a form of knowledge representation. I used the co-constructed member checked decision making timelines to understand the chronological order of events and identify critical incidents. Based on the transcripts and time line, "re-storyied" narratives (Creswell, 2007, p. 56) were generated. The reason I restoryied the narratives was because field texts from interviewees may not follow a story-like sequence of beginning, middle, and end (Creswell, 2007). In addition, a story could contain information not directly related to the event (Creswell, 2007), so this information was filtered out of the narrative. The narratives were reviewed by the interviewees to ensure that they were accurate. In addition, models were constructed based on the field texts and other knowledge representation artifacts.

Sweep Four. Sweep 4, called *deepening*, was conducted by asking "What if" type of questions to obtain a deep understanding of the expert's shared event and reasoning. The desired outcome was to obtain a "comprehensive, detailed, and contextual account of the incident" (Crandall et al., 2007, p. 78). Sweep 4 resulted in a deep understanding of the expert "decision maker's point of view" (p. 78).

Sweep Five. In Sweep 5, I studied the research participant's account and encouraged him or her to reflect on alternative outcomes either to the overall incident or to a particular critical instance that occurred during the account. "What if probes" (Crandall et al., 2007, p. 79) were utilized to clarify decisions or subtle clues as well as to provide details of background influences. My interrogation of the research participant and his or her account expanded my understanding of the research participant's recollection of the decision making process.

Sweep Six. In Sweep 6 (which is not included in Crandall's Knowledge Elicitation CDM Model), I generated a descriptive summary that provided the research participant with an opportunity to confirm, clarify, correct, and verify the generated narrative based on his or her account. It also provided the interviewee with an opportunity to ask more questions about the study.

Sweep Seven. Sweep 7 consisted of the identification of themes, generation of categories, and codes. The process utilized Creswell's (2007) narrative analysis procedure (see the section, Data Analysis, in this chapter). I conducted data analysis by analyzing the interview transcriptions, narratives, and critical incidents/decision timeline. I marked the interview transcripts with a highlighter and labeled sections with codes that emerged during data analysis (Krippendorff, 2004).

Sweep Eight. Content analysis involved placing words, phrases, sentences, and/or paragraphs into two categories: critical incident or decision. Then, I segmented relevant text passages. Coding small phrases in the transcription assisted in revealing the research participant's word choices. Coding longer phrases, sentences, or paragraphs assisted in understanding the context. To understand the evaluator's decision making, relationships between the two codes were determined, as well as the relationships between the critical

incident, the decision, and the evaluator's reason for the decision. The process aided me in identifying the factors that influenced the program evaluators' decisions and their decision making strategies.

Data analysis is an aspect of CTA. Creswell recommends narrative analysis "for capturing detailed stories or life experiences of a single life or the lives of a small number of individuals" (2007, p.55). Based on the research participant's accounts, also referred to as *field notes*, I generated a narrative. A description of the narrative analysis I used in this study is provided. In Figure 3.3, Creswell's recommended narrative analysis procedure that I modified to meet the goals of this study is depicted.

The narrative analysis followed this general procedure: (1) field notes were compiled from interviews; (2) I interviewed the research participant to understand the recalled account's background and re-story the account into a general framework; (3) a chronological framework or timeline was co-constructed; (4) the research participant and I identified causal links between critical incidents; (5) key elements and factors were explored; (6) I identified themes and codes based on revealed critical incidents, decision making strategies, and decision factors; (7) the research participant and I co-constructed a narrative; (8) my insight was interwoven into the narrative; (9) the research participant performed a member check of the narrative and timeline.

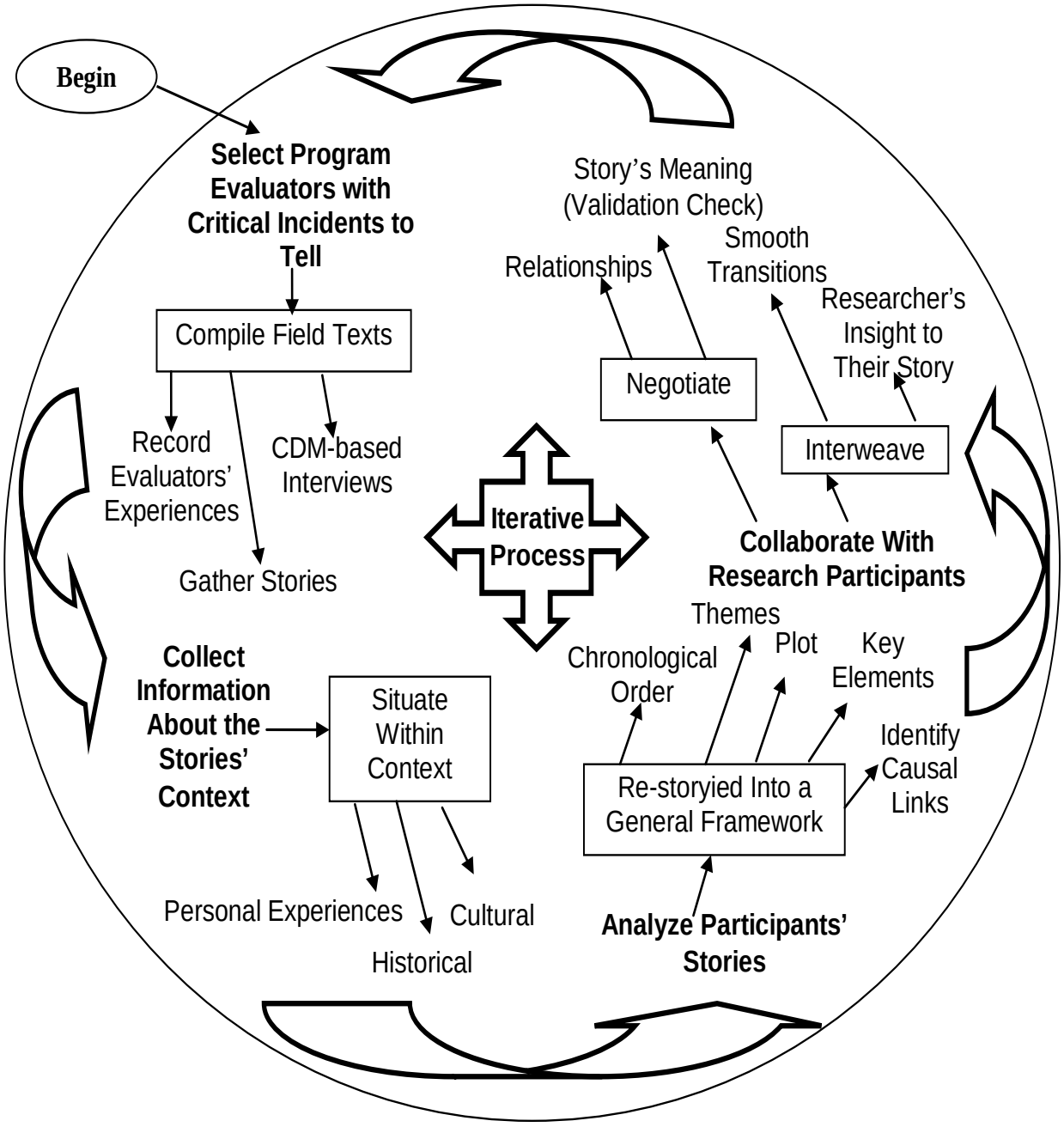

Figure 3.3. Narrative analysis procedure. Adapted from *Qualitative Inquiry and Research Design: Choosing among Five Approaches* (p.p. 55-57) by J. W. Creswell, 2007, Thousand Oaks, CA: Sage. Copyright 2007 by Sage Publications, Inc. Adapted with permission

In Table 3.2, a listing of the Sweep descriptions and the interview questions used in this study are provided.

Table 3.2

Sweep Descriptions and Interview Questions

Sweep Number	Sweep Description and Interview Question
Sweep 1	Introduce Research Study, Explain Consent Form, Background Information 1. How many years of experience do you have in evaluating HIV/AIDS health education programs? 2. How many of these programs were situated in sub-Saharan Africa? 3. How many of these programs were funded by international aid agencies?
Sweep 2	Incident Identification and Selection 1. Can you think of a time when your skills as a HIV/AIDS program evaluator were really challenged? 2. Can you share with me a time when your skills as an evaluator of HIV/AIDS programs really made a difference? 3. Can you share with me an instance when an HIV/AIDS program evaluation would have gone differently if you weren't there? 4. Can you tell me about the last time your decision making skills as an evaluator were challenged?
Sweep 3	Timeline Co-Construction, Verification and Decision Identification 1. Where in the time line should I put this? 2. Do I have this right? 3. Where in the timeline do you think this critical incident occurred? 4. When do you feel that this decision had to be made? 5. Where do you think that this decision resolved the critical incident?
Sweep 4	Deepening Understanding of Decision Making 1. What was the situation that let you know what was going to happen? 2. What was it about the situation that let you know what to do? 3. What led up to this decision? 4. What were your concerns at this point?

Sweep Number	Sweep Description and Interview Question
	5. How would you summarize the situation at this point? 6. What information did you use to make this decision? 7. How did you get this information? 8. What evaluation knowledge was necessary or helpful in this situation? 9. What were your specific evaluation goals at this time? 10. What were you hoping as an evaluator to accomplish at this point?
Sweep 5	Element Identification with "What if" Queries 1. What alternatives did you consider? 2. What would another evaluator possibly do differently? 3. What other actions could you have taken? 4. How would you have approached this decision earlier in your evaluation career? 5. How would this incident have turned out differently if someone without your level of expertise had not been there? 6. If you were to encounter this type of incident again, how would you approach it?
Sweep 6	Clarify and Validate Narrative - Conclude Interview 1. (While summarizing the critical incident story, timeline, and reviewing the key factors) What other observations would you like to add regarding your decision making process? 2. What other observations would you like to add regarding the critical incident(s)? 3. Did I misunderstand this? 4. Did I forget something? 5. Is there something you would like to add?

*Note Sweep Seven (7) is not included, because it refers to data analysis rather than the interviews.

Research Participants

On January 2010, an IRB application was submitted, and after the IRB application was approved, I began recruiting research participants. Through purposeful sampling, I recruited research participants who were expert evaluators with at least five years of evaluation experience. The programs that the expert evaluators evaluated were HIV/AIDS health education programs implemented in SSA. The research participants were identified through associations

and foundations such as the American Evaluation Association (AEA), International Program Development Evaluation Training (IPDET), Global Health Network, African Evaluation Association (AfrEA), UNAIDS **(**Joint United Nations Program on HIV/AIDS), Canada International Development Association (CIDA), CARE, World Vision, Oxfam, World Bank, International Development Fund, the Bill and Melinda Gates Foundation, and the Rockefeller Foundation. Most of the research participants were contacted through the membership's electronic media with permission of the membership president or public relations officer. Other research participants were evaluators whom I had met either through membership in evaluation associations or through participation in evaluation training programs.

Follow-up participant recruitment included contacting evaluators through phone calls and emails. In addition, I met with evaluators at evaluation conferences and global health conferences. The evaluators who were interested in participating in my research shared their resumes with me. I reviewed the evaluators' resumes to determine whether they had at least five years of evaluation experience and had conducted evaluations of HIV/AIDS health education programs in SSA.

Data Gathering

After identification of the criterion sample (Creswell, 2007), this qualitative research utilized audio-recorded, semi-structured retrospective interviews that were conducted through telecommunication. I transcribed the interviews. Chapter Five provides an in-depth description—based on Case One—of how the data were gathered for this research. The use of CTA, CDM, Narrative Analysis, and Phenomenography assisted in identifying the types of critical incidents that the expert evaluators encountered. It also helped identify the factors that influenced their naturalistic decision making, and how they made naturalistic decisions.

Data gathering was conducted over a four-month period, because (1) the research participants were very busy and (2) between the first and second interviews, the research participants were conducting evaluations in areas where telecommunication was limited for at least a month or more. Another aspect that added time between first and second interviews was that it took at least two to three weeks to generate rough drafts of the narratives and decision making timelines that were used for the second interview. In addition, after the research participants had received the rough drafts, they needed at least a week to read them before the second interview could commence.

Data Sources

My study utilized four data sources to create an in depth "story" and models of the evaluators' experiences. The data sources were the interviews, field notes, decision timelines, and narratives. Feedback was incorporated into the final report if it provided clarification or information that resulted in changes to the initial interpretation of the evaluator's experiences.

First interview. Semi-structured interviews began with open-ended questions, which according to Kvale (2007) are helpful in conducting an exploratory study. The questions used in my study deviated from the protocol, because they had to be adapted to each research participant's experience. All of the interviews were audio recorded. The interviews helped establish rapport with the research participants and provided the research participants an opportunity to discuss their evaluation experience. The first interview provided a forum in which a research participant selected an evaluation experience that met the following criteria: the evaluation of an HIV/AIDS health education program implemented in SSA; and as an evaluator, the research participant made decisions in response to the critical incidents they encountered.

Second interview and member check. The second interview concentrated on member checks, which served as a validation process. In this study, the drafts of the co-created decision making timetables and the narratives generated from field notes were shared with the research participants during the second interview. Research participants were invited to provide feedback on both the decision timeline and the narrative. This provided the research participants with the opportunity to confirmation or further explain (Yin, 1995) their experience. It also allowed for the comparison of data sources to "minimize misrepresentations and misunderstanding" (Yin, 1995, p. 109), "give grounds for revising the interpretation" (p. 110), and minimize distortion of the research participants' narratives.

Logistics. The interviews were conducted via communication media that best suited the interviewees' needs. In this study, the experts were "person[s], disposing of special knowledge and decision making power" (Littig, 2007) with at least five years of evaluation experience. One of the challenges often encountered when conducting research with experts is that their time is limited and valuable, so it can be difficult to arrange interview opportunities. What was encouraging was that, as Schank stated, experts' lives "can be full of stories" and people like to tell these stories about their experiences (1990, p. 111). The research participants seemed to want to share their experiences yet they also experienced many demands on their time. This circumstance extended the data-collection period. In some instances, many experts utilize gatekeepers to reduce interruptions that distract them from their work. This possible hindrance was not encountered in this study.

Challenges. Often, there are challenges related to dissimilarity between the person conducting the interview and the person interviewed (Littig, 2007). One of the concerns when conducting interviews is that the interviewer tends to have more power than the interviewee.

The fact that the interviewees in this study are experts in their field may have lessened this power imbalance and placed the interviewees and me on a more even level or it may have even skewed the power toward the expert. In addition, the expert may ask to view the interview questions and protocol before agreeing to the interview. Therefore, it is recommended that different coping strategies be used for gaining information. The strategies used in this study were to become an "insider" or being knowledgeable enough to be able to understand the expert's domain. Another strategy I used was to explain the research goals and required time commitment. Credibility was established by providing references to assure the interviewee that I was not wasting the expert's time.

Benefits and Risks

Some potential benefits and risks were anticipated for the research participants in this study. To address possible risks, the research participants were provided the opportunity to withdraw from the study at any time, for any reason. The benefits to the research participants included providing evaluators with an opportunity for self-discovery through reflection on their decision making processes and their approaches to solving challenging evaluation situations.

Evaluations of social programs are of benefit to society (Rossi, Lipsey, & Freeman, 2004), because they ideally inform and lead to program refinements. Stakeholders of HIV/AIDS health education programs are counting on evaluations to provide the necessary data to make evidence-based decisions regarding how to improve health education programs and thus, help people affected or infected by HIV/AIDS. Therefore, society benefits from understanding the critical incidents that evaluators face, their decision making processes, and the factors that influence expert evaluators' decision. Furthermore, by sharing their evaluation experiences, the evaluators can inform evaluation practice.

The possible risks to the research participants include that evaluators may share experiences that do not present them, the evaluation, or stakeholders, in the best light. To protect the evaluator, the evaluand and the stakeholders' personal information was not revealed in the story. For instance, the evaluators' name, race, and ethnicity were not revealed.

Limitations

The limitations of this type of research included the inability to generalize the research findings across domains to all program evaluators' decision making and non-participation by certain types of experts. In addition, as stated by Creswell, "the study of more than one case dilutes the overall analysis; the more cases an individual studies, the less the depth in any single case" (2007, p. 76). There are also concerns about the validity in qualitative research. In this study, a focus on establishing the credibility of the research was included to provide transparency via an audit trail. My research had the research participants validate their narratives and timeline to address the issue of "trustworthiness" in the study.

Non-participation in a study by certain kinds of experts represents another possible limitation. Experts tend to be extremely busy, and securing a first or second interview can be challenging. In addition, the research participants in my study were conducting evaluations in countries where they were hard to reach for the second interview. Every effort was made to not bias the selection of individual participants, but the very busiest (and perhaps most expert) evaluators in this context may have declined to participate simply because of time constraints. Cognitive science has limitations, one of which is that thoughts are self-reported. As of yet, there is no way to capture a person's thoughts without them being "expressed" in some manner of the written or spoken form. In addition, as stated by Hobbs (2003), there are problems when language is used to understand an event. The researcher

> should not lose sight of the fact that when this language is used, we are in the hypothetical world of explanation….[There is the] assumption that [interviewers' and interviewees']…judgments and other interpretation reports are reliable, privileged data…The second [assumption] is that such reports have something to do with the listener's interpretation procedures….[Thirdly, there is] the idealization of the "speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly" (Chomsky, 1965, p. 49).

There are several challenges to NDM, including concerns about the validity of the findings. To meet these challenges, models generated through research on NDM "have to describe what information decision makers seek, how they interpret it, and which decision rules they actually use" (Lipshitz, Klein, Orasanu, & Salas, 2000, p.10). Lipshitz et al. (2000) highlights that since "proficient decision making is driven by experience-tied knowledge … [it] puts a limit on the utility of abstract formal models" (p. 11), because this type of knowledge is (1) domain- and context specific (Ericsson & Lehman, 1996; Smith, 1997). As a result, many of the models cannot be generalized.

The critical decision method's heavy reliance on retrospective interviews was considered a limitation. Not all interviewees are adept at telling stories; therefore, the critical decision method placed a heavy cognitive load on the interviewee and me. The critical method was very time consuming, which posed a challenge when recruiting or interviewing experts. Crandall et al. (2007) recommends that the interviews be conducted by two interviewers. In my study, it was not possible to follow this recommendation, so this circumstance was a limitation in the knowledge elicitation process in particular. As a part of the validation and authentication process to mitigate the limitations of the knowledge elicitation process, I requested that the interviewee authenticate or confirm the accuracy of the narrative. All of the research participants reviewed and confirmed the accuracy of the narratives and decision making timelines.

Challenges were encountered, because I do not speak the languages of many of the research participants if the language was other than English. This condition means that I did not interview evaluators who did not speak English. This is a limitation, as in SSA, many languages are spoken, and not every evaluator speaks English.

Cultural challenges also needed to be taken into account when conducting this study. I have family members from SSA. However, it would be inappropriate to assume that I understand the cultures present in SSA. In addition, in this study, most of the evaluators were working for or with international development agencies. I have not worked for an international aid agency. Consequently, cues may have been missed or misinterpreted. It is hoped that co-construction and member checking helped reduce misinterpretation of the evaluators' experiences or the possibility of missed cues.

Technology such as Skype ® and telephony was used to contact expert evaluators. Some evaluators work where no communication technology infrastructure exists, or if one does exist, then it is unstable. This challenge was encountered when calls were dropped or internet connections failed. However, dropped calls or unstable internet connections did not cause the research participants to opt out of the study. In one case, due to a hurricane in the United States, the first interview was shortened, and the second interview was completed.

Validation

To address issues regarding researcher bias and increase the trustworthiness of the study, several validation procedures were utilized. Validation procedures such as member checking and an audit trail were used as a means of discovering subjectivity. Member checking was used to provide research participants with the opportunity to agree with, correct, or further clarify

knowledge representations derived from the interviews. Chapter five provides a description of the validation procedures.

Peshkin (1988) recommends introspection as a way of discovering subjectivity. Introspection includes being aware of wanting to go beyond the role of a researcher. Lincoln and Guba (1985) and Stake (1995) recommend the use of an audit trail. An audit trail was utilized—in particular, memoing, as a means of record keeping and encouraging reflexivity (Lincoln & Guba, 1985). As stated by Yin, the organization of the "chain of evidence" (1994, p. 78) (See Chapter 5 for more information) provides the outsider with the opportunity to follow the research design through to the conclusions.

Triangulation is another validation procedure. Data triangulation—the use of multiple data sources—was used in this study. Data sources included interview-based field notes, decision timelines, and narratives with thick, rich descriptions. Introspection revealed through the audit trail was another form of triangulation. The memos and code listings provided a record of the data analysis procedure.

Generalizability

As stated by Stake, "The real business of case study is particularization, not generalization.… There is an emphasis on uniqueness … the first emphasis is on understanding the case itself" (1995, p. 8). While the emphasis is on uniqueness, as recommended by Stake (1995), attempts were made to provide clear and easy-to-follow interpretations of the data. This effort will provide researchers with the opportunity to draw their own conclusions with regard to the generalization of the research findings.

Researcher's Perspective

Researchers such as Bogdan and Biklen (1992), Peshkin (1998), and Wolcott (1995) indicated that subjectivity or bias is natural. Bogdan and Biklen further indicated that immersion in the data and an investment of time help provide a detailed version of the event being studied (in this study, the evaluators' renditions of their experiences). Naming the researcher's biases increased transparency as well as provided an opportunity for reflexivity.

Biases, present when working with experts or those knowledgeable in their field, could skew the findings, because the research is likely informed from an elite perspective (Miles & Huberman, 1994). This may occur, because my study utilizes research participants who are experts and who are economically empowered to evaluate a program aimed at assisting developing countries in the midst of a health crisis. Many of the evaluators were working with people less empowered than they were and with people who were addressing or facing a humanitarian crisis in a very personal way. This affected the stories being told and the evaluator's perception of the event being studied. Attempts were made to obtain research participants who are evaluators from the country or regions in which the programs are being implemented. This was challenging, as the evaluation capacity in developing countries has not reached a critical mass.

Another potential bias was that I am a member of the evaluation community. While not at an expert level, I understand the challenges that evaluators may encounter. Every effort was made to avoid a complete loss of objectivity and perspective. A conscious effort was made to avoid becoming overly involved when establishing rapport with the research participants. It is important that my role as a researcher is clearly defined (Stake, 1995). The audit trail and data were monitored for critical language or a tendency to judge.

Furthermore, it should be noted that I am from a developed country and cannot fully understand the incidents that evaluators encounter when conducting a program evaluation in a developing country experiencing a humanitarian crisis. My experience with conditions and challenges while living in a developing country hopefully reduced the potential for bias and a lack of insight. I have lost several relatives to HIV/AIDS; therefore, it was important that the translation of these experiences into subjective judgments be avoided.

Finally, I have studied evaluation and decision making for more than three years. Familiarity with other researchers' work and reading HIV/AIDS health evaluations, white papers, and academic papers unavoidably influenced my understanding and interpretation of the research findings. Phenomenology requires the "bracketing" (Creswell, 2007, p. 59) of biases. An effort was made to discover biases and subjective responses.

Ethical Considerations

Qualitative researchers need to address ethical concerns. Several principles were addressed when conducting qualitative research: securing informed consent, protecting participants from harm, and ensuring confidentiality. In this study, the research participants were informed about the procedures of the study. They were informed about the number of sessions and their anticipated length of time. My goal was to ensure that research participants fully contributed to the study and did not stop participating in the study, because they were surprised at the amount of interaction needed. Research participants were informed of the opportunity to resign from participation in the study at any point, as well as the opportunity to take their data with them, if they so chose to. The consent form clearly stated the purpose of the research. In addition, the research participants had an opportunity to review the decision timelines as well as the narratives generated from their interviews so that protective changes could be made at their

request. The research participants' names were not used to protect their confidentiality both in the raw data and in the reports.

Pilot Study Findings

Analysis of the interviews in Fitzpatrick, et al. (2004) and relevant literature (see Chapter Two) indicated that the characteristics of NDM could be found in the interviews of evaluators' experiences in the field. Consequently, I conducted an IRB approved pilot study in July 2010 that involved the exploration of a program evaluator's decision making when facing critical incidents. The purpose of the pilot study was to answer the following question: How do program evaluators make decisions when facing critical incidents? The overall question is divided into two sub questions: What critical incidents did the evaluator face? What decisions did the evaluator make when faced with critical incidents?

For this pilot study, a participation request was sent to an evaluation organization based in the southeastern United States through the organization's listserv. An evaluator from the organization's health education evaluation special interest group responded to the participation request. The interview took place in the southeastern United States, at the time and location most convenient for the interview subject.

The participant was a female of 35 or more years of age who had completed a Master's degree in health education and promotion, as well as a Ph.D. in the social sciences. The participant had at least ten years of program evaluation experience. The participant was an evaluator of an HIV/AIDS health program that taught psychosocial skills to teenagers living with or affected by HIV/AIDS. A retrospective semi-structured interview was utilized to elicit a critical incident event and to reveal the evaluator's decision making process.

The interview was audio recorded and transcribed by the researcher. To stimulate discussions around evaluators' decision making, the Critical Decision Method (CDM) was used for "knowledge elicitation, data analysis, and knowledge representation" (Crandall, Klein, & Hoffman 2006, p. 9). The CDM was employed, because it has been found to be conducive to examining and analyzing decisions (Salas & Klein, 2001).

Data analysis was conducted by analyzing the interview transcription, narrative, and critical incidents/decision timeline. The findings from the study indicated the following aspects:

- the evaluator made her decisions through solution formulation based on a situation analysis;
- the evaluator utilized satisficing, which, as defined by Visser (2002), involves "looking for good, acceptable, satisfactory solutions rather than one and only best, optimal solution from all possible" (p. 3) solutions; and
- the evaluator did not generate a list of options but rather generated one option that she felt was "realistic."

The results also indicated that the evaluation was conducted under significant time constraints that resulted in real-time reactions, and the decisions made were in collaboration with others. The findings indicated that the lack of time had the greatest influence on the evaluator's decision making.

A decision making model was developed based on the pilot study (See Figure 3.4). The model depicts the evaluator's use of situation assessment and a reliance on requirement gathering and resource assessment. The decision making process of the evaluator was influenced by the following decision making factors: Time, Data, Human Resources, Knowledge, Experience, Technology, Motivation, and Ethics. In the pilot study, these were labeled as *decision filters*.

The evaluator's decision making process consisted of developing an initial decision, testing out the decision, explaining the decision, and negotiating with stakeholders. The decision was adapted based on the stakeholder's feedback. The adaption was based on a real-time response to the stakeholder's feedback. The research also indicated that the evaluator relied on satisficing to determine the decision solution. The number of tests, explanations, negotiations, feedbacks, and adapt cycles depended on the number of times that decisions were made to arrive at a final solution.

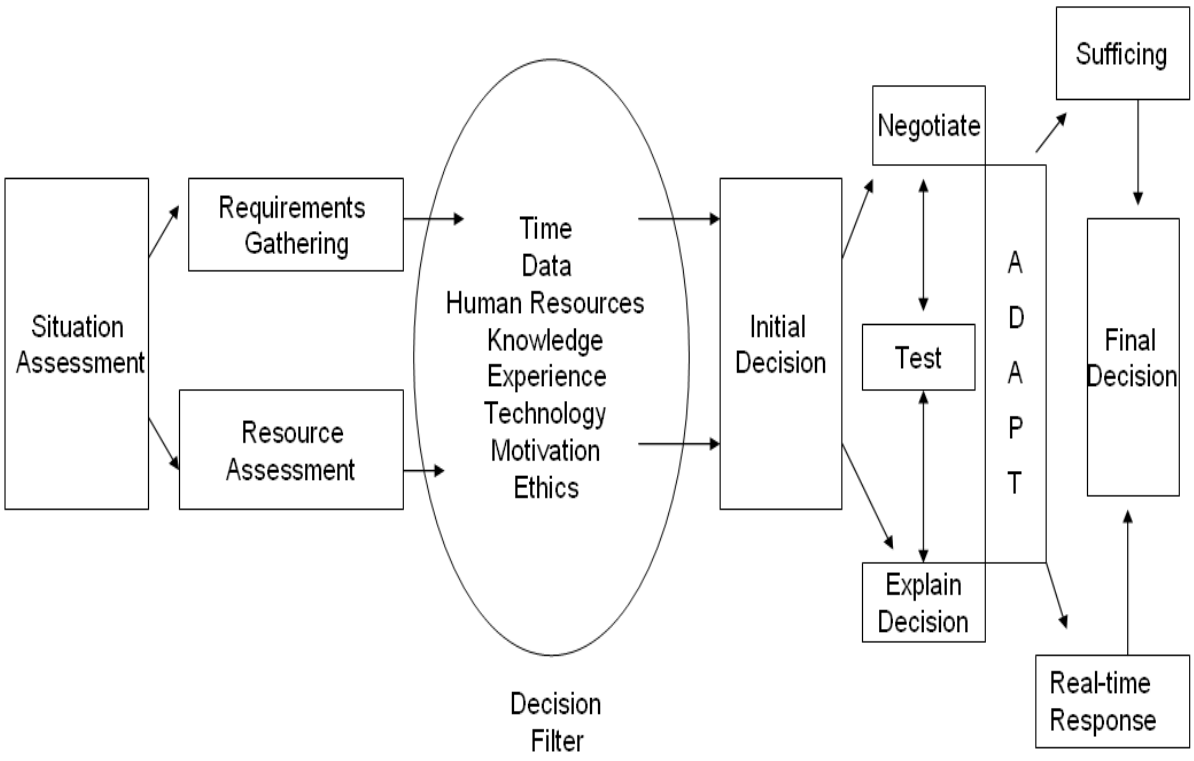

Figure 3.4. Pilot study: Expert evaluator's naturalistic decision making model

Chapter Summary

This chapter has described the study's process. To conduct my research, the critical decision method proposed by Crandall et al. (2006) was adapted. During data analysis, narrative

analysis with its inductive and dualistic approach was used. Phenomenographic methods were used to understand the variety of critical incidents the evaluators encountered and the factors that influenced their naturalistic decision making. This chapter also provides tables to identify the philosophical assumptions of qualitative research, how the assumptions were addressed in the study, and a matrix that describes the interview sweeps and lists the interview questions. A summary of the pilot study conducted in July 2010 is provided. Figures are provided that depict the three components of Critical Task Analysis, the research design, and narrative analysis.

CHAPTER 4

CASE DESCRIPTIONS

This chapter provides additional information about the case study method used in this research and a general overview of the seven cases that emerged from the analysis of the transcripts of the two one-hour, semi-structured, retrospective interviews I conducted with the participants. The descriptions focused on seven participants' stories based on their evaluations of HIV/AIDS health education programs implemented in Sub-Saharan Africa. The seven case descriptions were derived from member-checked, co-constructed narratives.

In case study research, one of the purposes is to understand the uniqueness of each case (Stake, 1995). To help identify and differentiate between the cases, case numbers and titles are provided. The cases are numbered one through seven. The case titles identify the evaluand and the type of evaluation that was conducted. This chapter also contains tables that describe the participants' demographics, types of evaluations they conducted, evaluation designs they utilized, types of data gathering instruments, a list of the development agencies that funded the evaluations, and a high-level summary of the cases.

Yin (2003) stated that there are several approaches to writing case studies. One such approach is "the type of multiple-case report [that] will contain multiple narratives" (p. 147). Due to the extended length of each of the narratives, the seven member-checked, co-constructed narratives were summarized into seven case descriptions. The purpose of the summaries was "to briefly describe episodes or to illustrate an aspect of the case" and to provide the reader with an increased "understanding of the case" (Stake, 1995, p. 128).

Similar to Stake's (1995) recommended organization for a case study report, the case summaries were designed to provide the reader with the following characteristics:

- the expert evaluator's profile and philosophical assumptions
- an opportunity to understand the evaluator's experience by providing background information about the case
- a description of the critical incidents that the evaluator encountered and the decisions he or she made
- a depiction of the complexity inherent in the evaluation experience
- details and quotations that assist in supporting the findings in the final report

Since each of the co-constructed, member checked narratives is at least 16 pages long, I decided to present the individual cases in a summary format that provides a "description one by one of several major components of the case" (Stake, 1995, p. 127). The case descriptions were divided into discrete sections detailing the expert evaluator's profile, evaluation case background, philosophical assumptions, critical incident(s), and decisions.

Marton (1988) stated that phenomenographic researchers are interested in mapping a person's experience and in developing a better understanding of a phenomenon and its environment. Phenomenographic research allows researchers to map each research participant's interpretation of an encounter with a phenomenon at several levels in which different understandings of the phenomenon are qualitatively depicted.

Narrative research attempts to provide a thick, rich description in the research participant's first or second voice. According to Creswell (2007), in narrative research, the researcher's voice is inevitably interwoven with the research participant's story. Keeping in mind the phenomenographic research's subjective approach, the narratives were member checked to verify that each research participant's voice was captured as accurately as possible.

Since the description of an evaluation's context includes providing demographic information (Greene, 2005), the evaluator's gender, current profession, academic training, residency, and years of evaluation experience are provided (an *evaluation expert* is defined in my study as an evaluator who has at least five years of evaluation experience). The role of the evaluator, defined as an external or internal evaluator, is included in the evaluator profile, since the evaluator's position in relation to the program or organization being evaluated can influence the goals and value the evaluator brings to a project (Patton, 2008). Patton (2008) indicated that it is important to understand the evaluator's responsibilities, so the evaluator's responsibilities are provided.

Goals are defined as the results that a person or system wants to achieve (Locke & Latham, 1990). As stated by Simon (1947), "Purpose provides the principal criterion in determining what things are to be done" (p.4). In evaluations, the goals of the evaluations, evaluators, and stakeholders guide the manner in which the evaluations are conducted. I decided it was important, therefore, to identify the goals present in the evaluation experience, as there was an interplay between goals, as the evaluator and the stakeholders worked "throughout the evaluation to focus the evaluation, participate in making design and method decisions, and to interpret results" (Patton, 2008, p. 72).

In cognitive task analysis, the goal is to reveal and describe the cognitive elements of goal generation, decision making, and judgments (Hoffman & Militello, 2009). In my study, a cognitive task analysis method, the Critical Decision Method, was used to reveal the research participants' fundamental assumptions about evaluation practice, their evaluation experiences, and to understand expert evaluators' naturalistic decision making. Martens (2009) indicated that belief systems are present in a research/evaluation context, and they influence how evaluations

are conducted. The expert evaluator's philosophical assumptions were described in each case overview, because (1) the evaluator's worldview had an influence on how the evaluation was conducted (Mertens, 2009); (2) phenomenography's subjectivist stance studies people's knowledge based on how they experience the world or phenomena in qualitatively different ways (Bowden, 2005; Marton & Booth, 1997); and (3) my study focused on expert evaluators' naturalistic or real-world decision making when conducting an evaluation.

In the case overviews, each evaluator's axiological (nature of ethics), epistemological and ontological (the nature of knowledge and the relationship between the evaluator, stakeholders, and reality), and methodological (approach to systematic inquiry) philosophical assumptions are provided. As much as possible, the philosophical assumptions were determined based on Merten's (2009) statement that

> [t]he axiological assumption asks the question: What is considered ethical or moral behavior? .... Ontologically speaking, how do we know that something is real … at a conceptual level …. Epistemologically speaking, we ask ourselves: What is the nature of knowledge and how do I come to know that the knowledge is 'true'? Is knowledge absolute or relative … how do I need to relate to those people from whom I am collecting data? .... Methodologically, I have choices that go beyond quantitative or qualitative or mixed methods to include how I collect the data about the reality of a thing in such a way that I can feel comfortable that I have indeed captured reality. (p. 45)

As stated by Patton (2008), "context matters" (p. 73). Greene (2005) stated that evaluation context refers to "the site, location, environment, or milieu for a given program (p. 13)." Greene further describes the context as having dimensions that include

> (1) the descriptive and demographic character of a setting … (2), the material and economic features of a setting…(3) the institutional and organizational climate in a setting … (4) interpersonal dimensions of a setting … (5) the political dynamics of a setting, and (6) the physical, geographic location of a setting. (p. 13)

Background information is provided to describe the "pre-existing evaluation bounds, organizational features, and project characteristics" (Alkin, 1985, p. 45). Providing background

information can assist the reader of the case overviews to situate an expert evaluator's stories within their context.

As indicated by Patton (2008), evaluators may encounter "intense fighting over goals and values" (p. 236) and may need to manage stakeholder conflicts or challenges throughout different phases in the evaluation life cycle. In this study, a *critical incident* was defined as an event that had an influence on how the evaluation was conducted or on its outcomes. Since my study was informed by the research participants' recalled critical incidents, descriptions of the critical incidents that the expert evaluators faced are provided.

In this study, decisions are defined as the "select[ion of] alternatives which are conducive to the achievement of the previously selected goals" (Simon, 1947, p. 4). When the evaluators shared their evaluation experiences, they also shared their decisions. They also shared how they made decisions and the factors that influenced their decisions. The decisions the evaluators made, the factors that influenced their decisions, and how their decisions are made are depicted in the narrative summaries.

As noted in Yin (2003), when conducting research, "there are some occasions when anonymity is necessary." The most common rationale for providing anonymity is when the case study is based on a controversial topic and "anonymity serves to protect the real case and its real participants" (p. 158). My study involved evaluators' experiences in the field where there may be "difficult circumstances under which many evaluations have to be conducted" (Bamberger et al, 2010, p. 19) and that may reflect poorly on an evaluator. Therefore, identifiable personal information was not included in case overviews and summaries.

Case One: Formative Evaluation of an HIV Health Department and an Interview of an Evaluation Department Official

Evaluator's Profile

In Case One, the evaluator is a male with more than ten years experience conducting evaluations in Sub-Saharan Africa. He has a doctorate in the social sciences and was a lecturer at an academic institution in SSA. He has conducted extensive research and evaluations on HIV/AIDS in Sub-Saharan Africa. He was hired by the government as an external evaluator to conduct an independent evaluation. The evaluation was commissioned as part of the donors' funding requirements and to inform the health department's strategic plan. The government-led intervention was funded by multiple international aid donors.

In this recalled evaluation experience, the evaluator's responsibilities included working with another member of the evaluation team to interview the government project manager in charge of HIV health programs. The evaluator scheduled a one-hour interview with a senior government program management official in charge of HIV health programs. The programs "included HIV heath education."

Evaluation Background

The evaluator and his evaluation team were hired to assist throughout the life cycle of an efficiency and effectiveness evaluation. The government — in response to a group of international aid agencies funding requirements — commissioned the external evaluation. The evaluator designed the evaluation, worked with the interview team to conduct interviews, analyzed the interview data, and wrote reports. The evaluation focused on programs implemented by the government's health department in Sub-Saharan Africa, which included HIV health education programs.

Evaluator's Philosophical Assumptions

In this evaluation experience, an open-ended interview was used to gather qualitative data that would inform the evaluation. Previous evaluation and interview experiences had taught the evaluator to expect an interviewee to (1) be very professional, which included being brief and providing the needed information for the evaluation, (2) try to "charm you" by continually stating that they feel that evaluations are important, and then talking incessantly without providing useful information, or (3) be "hesitant to participate, [which] may be due to the political context" or their concern that "you are going to get them in trouble with someone. [As a result,] they may give you very short answers or very general answers."

He did not anticipate encountering an openly hostile interviewee. He wondered why the person being interviewed was so resistant to evaluation. He wondered if it was due to the political environment or if it was due to poor communication exacerbated by a dysfunctional bureaucratic organization. He concluded that the interviewee was avoiding being evaluated.

The evaluator indicated that he had a responsibility as an evaluator and a member of an evaluation team to gather credible evidence that could inform the evaluation report. His goal was to gather data that informed the evaluation. He considered his responsibility as more than "parrot[ing] what [the official] said" and that his job was to "make judgments on what people say" and "make a recommendation that has an effect" on the delivery of HIV health services.

The evaluator was of the opinion that the official "needed to talk to me," since it was a part of the official's job requirement. He tried to secure the interviewee's cooperation, because the official was "accountable to the donors and the tax payers who funded the program." According to the evaluator, the political environment around HIV affected all the stakeholders involved in providing or receiving HIV health services. The evaluator considered the Health

Department “a mess,” highly bureaucratic, and uncoordinated. The evaluator reported that “due to the previous decade of unrest, they [the health department] were experiencing high levels of staff attrition.”

The evaluator’s impression was that his “interaction with the manager was affected because the manager was the person in charge of the HIV health department … [and because] “HIV matters.” He revealed his feelings when he said, “I wouldn’t have felt that way if it wasn’t HIV or if the program was doing an average or even a decent job. I think it was the political environment and the scale of the HIV/AIDS problem that influenced my reactions” to the hostile interviewee.

The evaluator also shared his views regarding the importance of time and timing when conducting an HIV evaluation as a practice. He revealed that in “HIV evaluation the time constraints are more pronounced, and you are at the mercy of participants’ responsiveness. Often you schedule an interview two to three weeks in advance. Then people don’t show up [and] cause you to fall behind schedule … You [also] have to work with several people and you have a lot of [data] … and you might not have a chance to go over the written material.” His experience as an evaluator has taught him that “poor planning from higher up [donors and program implementers]” participants are often “over evaluated” and “over researched.” Compounding the problem are “onerous” project requirements “with many lines of reporting.” This causes evaluation participants to be trapped in a “duplication of a lot of effort” where the “effort can increase exponentially based on how many funders you have and who they are.”

Critical Incident

Interview avoidance. The interview did not start well. The evaluator stated that at the beginning of the scheduled interview, the official abruptly stated, “Who are you? What do you

want? Who sent you? Why don't you go talk to other people?" The official said that there had been a review process six months ago. Then she said, "And now I am being reviewed *again.* Go read the previous review reports."

The evaluator did not respond angrily. Instead, the evaluator decided to share the purpose of the interview, which further infuriated the official. The official responded, "I am very busy and I have many demands on my time ... [The] HIV health program is very demanding…. I have a meeting(s) with the [government] minister … [and] with the media … I can't be interrupted by a low-level [evaluation] consultant."

The hostility continued. While she packed her briefcase and prepared to leave, the evaluator thought, "I need to figure out how to get the information." He asked himself, "How do you work with someone who has quite reasonable concerns [regarding the high demands on her time and being over evaluated]?" and "How can I navigate the power differences [between evaluator and evaluation participant]?" Then, he remembered a previous evaluation experience where an evaluator whom he considered an expert had conducted an interview with high-level officials and it went very well. He formulated his next option based on that experience.

Meanwhile, the evaluator also thought about "going over the official's head" but decided on another course of action. The evaluator decided to reframe the interviewee's response in a way that showed empathy and acknowledged the official's concerns. The evaluator stated that as he was contemplating his response to the hostile interviewee, he thought, "It is important that I don't abuse the authority that comes with my role as an evaluator."

The hostility continued. The evaluator made another decision. The evaluator decided to say to the official, "This [interview and evaluation] is an opportunity for an outside person to say what you want to them to say … If you think you have something important to say that hasn't

been heard, then go for it." The evaluator stated that his years of research on HIV/AIDS gave him the confidence to confront the official. He said, "I felt as if I had learned something" about the HIV pandemic and the services the department was providing. He concluded that the interviewee was avoiding being evaluated.

The evaluator considered himself "an easy-going and non-confrontational person, but when he gets irritated, there is a clear change in character." "He really hadn't planned to do it," but he closed his notebook and firmly said, "Look, we don't have to be here talking to you. Your office wanted to do the evaluation ...We are sorry that the department has been duplicating process. We were sent to talk to you. This is your one chance to give feedback on what you think is important for you to do your job … We are able to make recommendations [on your behalf], but if you don't want to make use of this opportunity … [I will report that] 'You don't have much to say.'" As the evaluator prepared to leave, he emphasized, "It is up to you." The evaluator indicated that if this had been a research project, then he would not have been so assertive. He knew that the official was required to participate in the evaluation process, because "It was part of her job." The official decided to participate in the interview. Two hours later, the official was still talking.

Evaluator's Decisions

The evaluator made several decisions while the critical incident was occurring. His goal was to gather the information he needed for the interview report, which would be used to inform the evaluation. First, the evaluator decided to explain the purpose of the evaluation. When that did not work, he decided to reframe the problem. When that approach was not effective and the evaluator's level of frustration had reached a tipping point, he "decided to let the official know

that her lack of participation in the evaluation wasn't a good idea." He further emphasized that she was missing an opportunity to have some of her interests or needs met.

Case Two: Summative Impact Evaluation of an HIV/AIDS Health Education Program's Consortium Model

Evaluator's Profile

In Case Two, the evaluator is a female with over 19 years of experience in conducting evaluations in Sub-Saharan Africa. She was a consultant who had lived in Sub-Saharan Africa for almost ten years and conducted evaluations in more than ten countries in this area. She is a noted facilitator and a monitoring and evaluation specialist who holds a PhD in evaluation/international development and a Master's of Public Administration. She has done research that focuses on gender and evaluation. The evaluator in this scenario was trying to secure the donor agency's engagement in the evaluation design and gain the lead consortium's cooperation in the data gathering process. The evaluator was trying to gather the data she needed so that she could write up an evaluation report based on credible evidence.

Evaluation Background

The evaluator was hired by a foundation, which worked in collaboration with other international aid agencies, to conduct an external evaluation to determine the impact of the use of a consortium model on the HIV/AIDS health education programs they were funding. The evaluator was the recipient of the small contract that was to include five days in the field. Due to the small budget, the consortium leader was to provide her with transportation and secure a place for her to stay. They were also supposed to make all the arrangements for the interviews and focus groups. Her repeated attempts to contact the evaluation sponsor were unsuccessful.

Evaluator's Philosophical Assumptions

The evaluator made repeated attempts to contact the evaluation funder/intervention donor before she went into the field to gather data. By contacting the foundation funding the evaluation, she hoped "to understand what the donor wanted her to do, the evaluation questions they wanted answered, how to gather the data needed for the evaluation, [identify] who would be the evaluation participants, and how they [the donors] were going to use the evaluation." She knew that she "needed more information so she could design the evaluation and determine how to gather the information she needed for the evaluation." The evaluator is "heavily influenced by Michael Patton [Patton's theory on the importance of evaluation use], so I always try to ask [questions such as], 'Who are the users? How are you going to use the evaluation?'" The evaluator believed that evaluations are supposed to be useful to the donors and to empower the people who are the donors' clients.

The evaluator wanted to make sure that the evaluation was useful to the consortium members. Her belief that evaluations should help others "guide[s] my [her] decisions in the field." She said, "It isn't only because I am getting paid [for my time and expertise], but it's also because it is taking up participants' and my time. For example, [in the shared evaluation experience] the women that agreed to participate in the focus groups were taking time out of their schedules to work with me."

The evaluator also said, "I'm very big on providing as much capacity building as I can during evaluations … I work at trying to explain what evaluation is, how it should be working, how we could do it." Her approach to evaluation is guided by Michael Patton's concept of "process use." The first benefit of her approach was that the consortium members were able to express their concerns and felt that they were heard. The second benefit was that the consortium

members were able to understand that evaluation can be "helpful and useful to them—[it is] not just the donor coming out or sending someone out for five days." She made clear to the consortium members "that this 'evaluation' was not an actual evaluation."

The evaluator considers herself a "development worker" who happens to conduct evaluations for the betterment of the people in developing countries. She believed that "Even this experience [which was stressful] had some useful stuff come out of it. We did manage to run one focus group with members of the consortium, and they said that this was the first time they had someone come out, sit down with them, and listen to them talk."

Critical Incidents

Non-communicative evaluation funder. "It's like everything that could go wrong went wrong…. I tried to call [the evaluation funding agency] before I went out to the field because people define 'impact' differently … [and] I needed to know 'What they were trying to find out?' I can focus evaluation questions if they tell me what they want, why they are doing it, and who is going to use it. [That way] I can say [to myself], 'Look, I know the donor really wants to know X, so I need to interview those people.'"

The evaluator tried to determine why the donor would not return her calls. She wondered whether it was because the "donor requesting this evaluation was busy." She stated, "I did try to make [engagement] happen. I guess I could have tried to engage with them a bit more, but I doubt it would have helped."

Since the donor would not return her calls, the evaluator "needed to make assumptions and decisions in the field about what data to go after … [and] make assumptions about the focus of the evaluation." Due to missing and vital information, the evaluator used an "exploratory approach." The evaluator expressed her concerns about her decisions. She said that given the

situation, and "regardless of how university professors might frown on the methodology that I ended up using, I had to stay [to complete the evaluation]." The evaluator considered the "donor's lack of engagement . . . . a barrier to implementing an evaluation that was useful."

Evaluation avoidance. The evaluator said that a member of the lead consortium's staff was "supposed to fetch me at the airport, and they were supposed to book somewhere for me to stay. When I arrived at the airport, there was nobody there. I couldn't reach them on their cell phones." The evaluator went to the lead consortium's office and demanded to know, "Where were you? You were supposed to pick me up today!" They replied, "Oh, we thought it was next week." She thought, "They knew darn well that I was coming, but they also knew that they hadn't done anything. There is no way they forgot; there were too many emails back and forth about the dates, the plane ticket—everything." The evaluator stated, "I know it is my assumption, but I think they were avoiding getting evaluated."

The evaluator only had five days to gather data in the field, and the lead consortium staff had many problems securing the interviews and focus groups. The evaluator sighed, "If only the [staff at the] lead consortium had set up interviews with the other organizations, then I could have looked at evaluating the program design … [or] I could have looked at the results level." Eventually, a few interviews were scheduled. The evaluator said that she "used the interviews as an opportunity to listen to the consortium members' complaints."

The evaluator was able to identify that the other members of the consortium were not getting paid and that they were having difficulty working with the consortium leader. Over the years, she has found that "people in the field see me as the voice that goes back to the donor. They say, 'We want our money' or 'We need to be funded.' It has happened before with other AIDS programs."

Resistance to evaluation findings and report. The evaluator "wrote the evaluation report ... [A] lot of it [the report] was about [the donor's] theory of change, about their whole idea of having a consortium, with results indicating that it was really either not being [sic] the way to go or it was not being [sic] implemented correctly, and because of that, it was impossible for me, in that context, to find out about any kind of impact."

"The donor wrote back to me. They were quite angry … [They] said, 'This is not an impact evaluation. This is a site visit report.'" The evaluator stated, "I thought to myself, 'I can write you an impact evaluation, but it would be one paragraph long. There is no impact.'" The evaluator said, "I had evidence that supported my findings. I tried to explain it to them [that there was no impact] but they just didn't seem to get it."

Evaluator's Decisions

The evaluator's decisions included using an "exploratory approach," making the "evaluation useful," and using the process of conducting an evaluation for evaluation capacity building. The decisions followed a general pattern in which she asked a question and was not provided feedback; then, the evaluator made decisions based on her own experience, training, and judgment.

Case Three: Summative Impact Evaluation of a Micro-Credit and an HIV/AIDS Health Education Intervention

Evaluator's Profile

The evaluator in Case Three is a male with over 12 years of experience in monitoring and evaluation of HIV/AIDS health programs. He lives in Sub-Saharan Africa. He is a research and program advisor for an international aid agency whose offices are located in Sub-Saharan Africa. The evaluator's Master's in Health Systems focused on International Health and the HIV/AIDS

epidemic (epidemiology) in Sub-Saharan Africa. His employer provides support to many agencies implementing HIV/AIDS programs.

In this scenario, the evaluator and his team conducted an internal impact evaluation of an intervention implemented by one of his employer's many partners. In the evaluation experience that the evaluator shared, the implementation partner was hired to gather data/credible evidence that would inform the evaluation report's findings and recommendations. The evaluators collaborated with the NGO (non-governmental organization) and provided technical support to the evaluation implementation team.

Evaluation Background

The evaluator stated, "We had a partner, an NGO, that we were funding to do an intervention with young women…. The NGO's HIV/AIDS prevention intervention provided young women with micro-credit so that the young women could start their businesses, and provided them with HIV/AIDS health education. The goal was to reduce risky sexual behaviors."

The evaluator stated that "there were several of us: the prevention advisor, the NGO and me, the monitoring and evaluation advisor for the organization funding the evaluation. We got together and we decided to try to design an evaluation to see whether or not the intervention had an impact." The evaluation team "decided to measure the following risk factors for HIV transmission: changes in behavior, transactional sex, gender-based violence with young girls, number of sex partners, and inter-generational sex." The "[evaluation team] also planned to gather baseline measurements and compare them to follow-up measurements from another district. The other district would receive the HIV/AIDS health education component and then receive the micro-credit after the evaluation was done."

Evaluator's Philosophical Assumptions

The evaluator stated that the evaluation team was interested in conducting an evaluation of an innovative HIV/AIDS prevention intervention being implemented by an NGO, because research has shown "that poverty is a risk factor for HIV and AIDS because poverty can lead young girls to engage in transactional sex to raise money." The intervention provided young women with micro-credit and HIV/AIDS health education as a way to reduce risky sexual behaviors.

The evaluator believed that "when you do a novel intervention, it is the kind of thing that needs to be evaluated." He stated, "I think it was appropriate to conduct the evaluation even though we didn't have a lot of experience and literature to draw on. There had been a couple of attempts at evaluating similar types of interventions. One attempt didn't show much of an effect, and the other one kind of fell apart."

The evaluator stated that "the potential for bias is [was] a concern." The evaluator's previous experience indicated that "reported risk behaviors tend to go down just because of the introduction of the intervention. So you wonder if the changes are true or not." When conducting an evaluation, the evaluator asks himself, "Is it [the changes in risk behaviors] because you have been talking to them [evaluation participants] about risky sexual behaviors?" or is it that "they [evaluation participants] may want to give you information that they think you want to hear."

Due to a small budget, the evaluation team hired the NGO to conduct the evaluation. The evaluator acknowledged that it "was not good [practice] to have the NGO do the evaluations because they don't [didn't] have the [necessary] research expertise." The evaluator stated, "The NGO was very confident—you know, 'Oh, we can do this.'"

The evaluator said, “One thing that you learn very quickly is that partners will promise, whether they can deliver it or not.” The evaluator stated that “since NGOs are trying to get money for their projects … I think there is a stronger potential for bias [when an NGO is conducting an internal evaluation] than when the donor is evaluating their own project.” The evaluator found that “—just from hearing people talk—it is good practice to have an external evaluator.”

The evaluator highlighted the importance of field experience when conducting research when he said that “[the NGO] hired a PhD student to do the evaluation. She was very good and very smart. She had *not* done a lot of research in Africa and certainly *not* qualitative research.” The evaluator stated that “The problem was that a lot of her ideas about interviewing techniques and how to ask the questions were straight out of grad school…. If you try to conduct an interview [in the field] based on what you have learned in graduate school,—well, in the field, it comes off a little bit stiff and it isn't particularly effective…. You have to play with the rules a little bit in order to establish rapport.”

The evaluator acknowledged that the evaluation had many problems because of his inexperience and “bad judgment on the whole thing.” He stated, “When you have more experience, you can tell that the partner can't do the evaluation on this small budget. It wasn't the partner's fault. We were the ones who suggested the evaluation. It was just a lot of people that didn't have that much experience trying to do something that was a lot more complicated then we had anticipated.”

Critical Incidents

NGO’s credibility. The evaluation team “originally wanted to have a strong research design with some kind of randomized control comparisons.” The evaluator stated that “because

the intervention had started and the NGO had already chosen the micro-credit beneficiaries, we couldn't do a randomized evaluation … it turned out that "[the NGO] had made promises as to who was going to get the micro-credit. The NGO was really adamant that going back on their promises would really damage their credibility in the community." The evaluation team "brainstormed on what else we could do." Since "the same program was expected to be rolled out in another district," the team modified the evaluation design to a non-randomized "comparison group" evaluation design.

Lack of research expertise. The evaluator stated that "since the budget was so small that we [the evaluation team] couldn't afford to hire a high-powered research group [so], what we did was ask the NGO to do the evaluation." The evaluator stated that "the NGO was good at implementing the intervention, but they just didn't have the research expertise necessary to gather data that was of good quality…. It turned out to be a huge disaster." Compounding the problem was that "the people we [the evaluation team] hired to do some data analysis said that they couldn't link the baseline to the follow-up information. We couldn't link the sexual behaviors data to the economic loan data. That was a huge disaster. We took a look at the sexual behavior baseline data, and there was a 40% non-response to some of the behavioral questions…. You can't really use variables like that." What made the situation worse was that "for the micro-credit part of the intervention—the NGO had these extensive economic interviews and transactional behavior interviews—[the statisticians] weren't able to link [the data to] any changes in behavior."

Adding to the evaluator's concern about the quality of the data was that "because of the small budget, [the evaluation team] couldn't gather biological markers. Gathering that type of information can be very expensive. The biological markers could have provided a way to

compare or to support the self-reported risk-behaviors. This would help weed out self-reported bias."

Insufficient budget. The evaluator stated, "We needed a bigger budget so that we could have hired interviewers that had been properly trained and get more than self-reported risk behaviors like some biological markers, like STI's (Sexually Transmitted Infections). We also needed a bigger budget so we could increase our sample size. We had a pretty small sample size of about 700-800 people." Due to the quantitative study's poor data quality, that portion of the data was not used. The qualitative data were used for the evaluation report, because that data was considered to be of "reasonable quality."

Bureaucracy and miscommunication. The evaluation team submitted a conceptual proposal of the evaluation to their organization's headquarters, and they "thought that the evaluation proposal had been approved [so the evaluation team started implementing the evaluation]." The evaluator said, "Well, it turned out that there was a bureaucratic mix up. We sent in our original concept proposal to the evaluation funding organization's headquarters. Apparently, they have expert researchers that review proposals, and the reviewers were supposed to have sent us comments. We never got the reviewers' comments. We weren't supposed to start the evaluation until we submitted a full proposal with a protocol, until headquarters approved the protocol, and [until we had] gotten headquarters' technical expertise involved in addition to their final approval. We had a protocol that had been approved in the country office, but it turned out that it wasn't approved by headquarters."

The evaluator stated that "headquarters found out about it [that we had begun implementing the evaluation]. They said, 'How could your study have started without our approval of the protocol?' So they made us shut down the comparison study portion of the

quantitative study." The evaluator stated that "the quantitative survey we were conducting was allowed to continue" but it "was of poor quality and not very useful." The evaluator stated that "[the evaluation team] didn't need headquarters' approval for the qualitative design"; so, "we were able to continue with the qualitative study to find out, 'What impact did this loan have on the young girls and their risk behaviors?'"

Evaluator's Decisions

The evaluation team decided to modify the research design. They had originally tried to use a randomized control research design, but since the NGO had already made a promise regarding where they would implement the intervention, the team needed to compromise and use a comparison group instead. Due to budgetary constraints, the evaluation team decided to hire the NGO to conduct the evaluation data-gathering process.

The data turned out to be of low quality with a high non-response rate to interview/survey questions. The statisticians who were hired to analyze the data were not able to link baseline data to the time series data. The evaluation team's budget did not provide the team with an opportunity to gather biological markers so that they could triangulate the data. The team was worried about potential self-reported bias. Due to miscommunication, bureaucracy, and team inexperience, portions of the evaluation were defunded and halted. They had to settle for using a "qualitative end-of-project evaluation that was of reasonable quality."

Case Four: Ex-Ante Evaluation That Was Aimed at Designing a Transactional Sex and Women's HIV/AIDS Vulnerability Evaluation

Evaluator's Profile

The evaluator in case four was a female with more than 12 years of experience conducting evaluations. She lives in North America. She is working as an executive for an

international aid agency. Her work focuses on research and evaluation aimed at improving the quality of programs that focus on public health and the social and environmental factors that influence health. She has worked extensively both domestically and internationally while conducting research and evaluations that influence public health policies and practices to prevent HIV/STD (sexually transmitted diseases) transmission. She is a noted evaluation specialist who holds a PhD and a Master's in Public Health. The evaluator in this scenario was trying to secure "community entry" [entry into the community] in two countries so that the team could conduct an ex-ante evaluation.

Evaluation Background

The evaluator and her evaluation team applied for internal funding to conduct an ex-ante evaluation in two countries. The evaluation was to be designed to assist in the development of an intervention whose goal was to understand how heterosexual transactional sexual relationships are initiated or sustained (for material gain) and how they are related to HIV/AIDS risk/vulnerability among women. This intervention was to provide "HIV awareness, psycho-social support services, and information on abstaining from unsafe sex practices."

The evaluator stated, "the informative [ex ante] assessment would also help us design the intervention's educational component." The evaluator stated that in this study, they defined transactional sexual relationships as sex exchanged for a gift such as material assistance or service. The evaluator stated that in this evaluation, they were not studying commercial sex workers.

Evaluator's Philosophical Assumptions

"Before we [the evaluation team] could officially submit our proposal, we needed to make sure that we followed our own organization's internal requirements, such as writing a

protocol, which was a major endeavor." "Multi-country protocols can be quite laborious because you need to work with both countries." The evaluator stated that to "conduct a successful evaluation you need to gain an understanding of the culture, context, and the changes in the context [contextual changes] during the evaluation." The evaluator said that it was "a basic fundamental evaluation principle … [that] has been verified by my field experience … [and] what I have read about evaluation and the social behavioral sciences."

The evaluation team hired interviewers and observers who could speak the two countries' local languages and who were from the surrounding community. The evaluator stated, "I think it's important that people can relate to the interviewers rather than having a white person from the United States coming in to conduct the interview."

The evaluator reported that "in Sub-Saharan Africa, I observed other studies that didn't recognize the importance of planning entry into the community. I have seen the challenges that have occurred when evaluators did not take time to understand the context—what it takes to get a study underway."

The evaluation team "tried to have maximum privacy when conducting a study so that we can assure both participants' and data confidentiality." The evaluator added that "protecting participants' confidentiality is one of our principal responsibilities, so we educate participants about the consent process. It is also critical that we let them know why the research is important. Their increased understanding about the consent process and about research or evaluations can benefit their community and inform the design of future health programs." The evaluator continued to say "engaging and involving persons in the research [evaluation] is a part of the education process."

The evaluator explained that the organization she works for is "required to make sure that we are aligned with [the country's] mission." The evaluator said, "[it was] an important part of the process that you need to plan for. Sometimes it goes very quickly; sometimes it is slower than you would like. Sometimes there are competing priorities in the ministry, or the key people that you needed to talk to are not available…. [In addition,] we needed to keep in mind that the different countries had different processes, so for each country we needed to go through different steps to meet with the people from the ministry."

In this evaluation, the team used a geographic information system (GIS) as "an important tool and method to gather data." The evaluator stated, "[that] is not normally done." The team "also used methods such [as] in-depth interviews to gather data" and "used a methodological approach called rapid assessment to collect the data quickly." The evaluator stated that for the "next phase of the evaluation, [the evaluation team] will bring in an expert to help us design materials and tools that will be used in the intervention."

The evaluator stated that the evaluation required collaboration, so "communication was very important." The evaluator also stated that documentation of the team's "decisions was [sic] another part of the communication process" and the documentation became "a record of what you have agreed to do. That way you can make sure that you all know your protocol ... [and] the documentation can also be used to understand what is occurring during implementation…. [I]t can also be used when you are thinking about how to write the findings."

The evaluator stated that "evaluation takes patience and flexibility ... Some things you can anticipate, but a lot of them you can't. And then we have to rethink the timing and the scheduling. It can be very difficult to do that." During the development of the protocol, "it became clear, through the literature review and through subsequent discussions, that we needed a

lot more information on the women's economic, social, and sexual networks. That required us to expand that piece." The evaluator also stated that the team "found that the study was disproportionately qualitative. Our agency accepts qualitative work, but it is not as highly regarded." The use of GIS data "made sure that we had the balance of data so that we could substantiate and understand both (the qualitative and quantitative data)."

Critical Incidents

Problems, rumors, or miscommunications. The evaluator stated that when conducting an evaluation, "we want to make sure that the stakeholders are truly kept involved and informed. We want to stay on top of any potential problems, rumors, or miscommunications." The evaluator stated that her "years of experience working in domestic and international research [evaluation] have taught me to be proactive in planning community engagement and community entry." The evaluator stated that "just like any other type of research, you need to try to schedule a time that is convenient for the participant. We try to minimize any disruptions to their schedules. We try to reduce barriers to the facilitators working in the field and the evaluation process."

Securing ministries' support. The evaluation team decided to talk "to the ministries in each of the countries and all of the implementing partners." She stated that "there have been occasions where getting the health departments on board was challenging. It isn't that they are resistant to our meeting with them; the challenging part was that we needed to sit with them and explain how our proposal relates to their broader mission." We also asked them to help us "identify the populations or country context they were most interested in working with and exploring." The evaluator believed that "it was important that we worked with each of the countries' key stakeholders such as the Ministry of Health … key community leaders ... [and] the

potential participants in the study." The evaluator stated that "when we go to the country [and during the structured observations] we have a checklist of items that we want to go through."

Conflicting previous or ongoing research. The evaluator stated that "when we got to the country, we try to determine that there aren't other duplicative types of research going on … and if there was other research in the community that could cause challenges. We also tried to determine what previous research has been conducted. So, when we talked to people and asked them about a study or research, we asked ourselves 'Do they know what those words mean?'"

Insufficient resources. During the observations, we tried to look at what was available, such as "Is the Internet available in the community?" We also tried to determine whether we had staff who could support the project.

Conflicting agricultural or weather cycles. The evaluator also stated what we tried to determine: "'What are the agricultural cycles?' We looked for cycles [such as seasons or weather patterns] that we need to be aware of."

Work, residence, and migration patterns. The evaluator stated that the team hired people to interview "the women in the population that we were studying. We wanted to understand what was occurring where they work, live, and travel. The information gathered was used to inform the team as to when our presence would not interfere with the women's work and their businesses. In some cases, we interviewed or identified people that could inform us of the things we need to be aware. For example, we needed to know what were especially busy times or when it's not convenient to talk to people or whether there are events that make it more difficult to gain access to people." The evaluator stated this was part of their general protocol.

Two countries and their populations. The evaluator stated that for this evaluation, the team conducted "a thorough literature review" for each of the countries. She also stated that the

team decided to visit "the two countries at least three times before we began writing the protocol" so that they could understand the evaluation context. The evaluator stated that as a part of the official protocol, "the team conducted structured observations to understand the evaluand context and the population being studied." The evaluator stated that "[the evaluation team's] decision to conduct structured observations is supported by the literature" and by knowing "what observations can bring to the table."

Coordinating with multiple stakeholders. The evaluator stated that "a major evaluation challenge that we have encountered when conducting evaluations in Sub-Saharan Africa is that there are often multiple parties involved in an intervention. That means we needed to coordinate with all of the parties involved in an intervention. This included stakeholders and staff at our central office, field offices, the community leaders, interested implementation partners, and government ministries for each of the countries."

The evaluator stated that they decided to bring in "collaborators with different types of backgrounds. We have an anthropologist, a sociologist, people that specialize in GIS, and a network specialist/analyst." The evaluator felt that "bringing in a diverse range of people strengthens what you can do and brings in different perspectives."

The evaluator stated that the project required "a lot of collaboration. We had ongoing weekly calls. Sometimes it was more than weekly. How often we collaborated depended on the phase of the evaluation's life cycle." The evaluator stated that documentation was used to "determine if the protocol is being implemented as planned because if you deviate from the plan, you are in violation of the IRB." The documentation can be used to help thoroughly describe the study and how it was developed, designed, and implemented.

Rapid data gathering and need to track trends. The evaluator stated that they needed to gather the data quickly and track trends. The evaluator stated, "We have used very simple mapping tools before. This was the first time we utilized tools such as a GIS system to measure coordinates and identify specific trends. The data gained from the structured observations, from the information collected directly from the participants, and from the use of different base maps will be entered into a GIS system. By using base maps that contain information about transportation routes and using the information we are gathering, we are able to, for example, identify key transportation routes and trends." The evaluator stated that the data stored would enable the team "to map and identify key components of the environment or context that may influence or impinge on these women's behavior."

Language and culture. The evaluator stated that in one country, they "established entry into the community" by selecting interviewers "from the country and proximal areas. In one community, the interviewers were actually from the community we were studying. We wanted people that spoke English and at least one of the two local languages spoken in that country. It was critical that we hired interviewers from the surrounding areas that could speak the country's indigenous languages."

Bureaucracy and non-compliance. The evaluator stated that "eventually funding was approved, but disbursement of funds is always delayed by a year since there is always a gap between getting the idea submitted, approved, and actually receiving the funds." The evaluator stated that they "had to defund the evaluation because the award recipient was not doing what they were supposed to do. We had to re-announce the funding opportunity and select a new award recipient and award the funds to the selected applicant." The evaluator stated that it "added six months to the implementation timeline."

Civil unrest. The evaluator stated, "We started the formative evaluation in one of the countries. Unfortunately, the evaluation had to be put on hold due to civil unrest. Hopefully, we will be able to go back in the end of June."

Ruling against human subject research. In the other country where the formative evaluation had been postponed, just about the time when we were ready to start the evaluation, the Ministry made a ruling that there was no basis on which to do human subject research in the country at all. Just in case we were not able to conduct the research in country two, we pursued alternative study sites. I consider that an unexpected barrier to community entry.

We kept in contact with the key stakeholders and, in particular, with the communities we would be working in, because we had said, "'We are coming' and then we didn't come. We had to go back and explain that there were challenges that needed to be dealt with before we could appropriately come and speak with them." The evaluator stated that "at first it seemed like it would be a two-month delay due to the Ministry's decision, but there were further delays because of weather conditions."

Evaluator's Decisions

In order to assist in securing community entry, the team decided to talk "to the ministries in each of the countries and all of the implementing partners." The team decided to conduct "a thorough literature review" for each of the countries and make two field visits before they wrote the protocol to understand the two countries' evaluation contexts.

To identify potential evaluation barriers, the evaluation team conducted a resource assessment. They also conducted research to determine trends, understand previous or ongoing evaluations, ascertain whether potential participants could understand terms such as *evaluation* or *research,* and determine the agricultural cycles and weather patterns. The evaluator stated that

for this evaluation, the team conducted “a thorough literature review” and “conducted structured observations to understand the evaluand context and the population being studied.” The team also hired people to interview “the women in the population that we were studying … to understand what was occurring where they work, live, and travel.”

The evaluator also stated that in order to collaborate, they coordinated ongoing weekly calls with the stakeholders. The evaluators also used documentation to “determine if the protocol is being implemented as planned because if you deviate from the plan, you are in violation of the IRB.” Since the initial award recipient was non-performing, the evaluation team had to de-fund the award, re-announce the funding opportunity announcement, select a new award recipient, and award the funds to the new selected applicant. Eventually, it was decided to place the evaluations on hold due to civil unrest in one country and because in the other country, the Ministry had made a ruling that prevented human subject research. The evaluation team kept in contact with the key stakeholders to keep them informed of the situation.

Case Five: Summative Evaluation of a Secondary-School HIV Health Education Intervention

Evaluator’s Profile

In Case Five, the evaluator was a man with more than 30 years of international experience in conducting evaluations. He lives in Asia. He has worked in more than 50 countries His work involves conducting research and evaluations with a focus on the educational response to HIV and AIDS. He holds a Master’s of Arts in English as a Second Language, a PGCE (Postgraduate Certificate of Education) in the social sciences, a BSc (Bachelor of Science) in Sociology, and a PRA (Participatory Rural Appraisal) in Adult Literacy.

He is a noted evaluation specialist who has done extensive work on the challenges that teachers face when addressing HIV in their work. He is currently an independent consultant conducting evaluations. In this evaluation experience, the evaluator was not working as a consultant but was an "employee of an HIV/AIDS government-based development funding agency." The evaluator stated that he "was brought in [as an evaluation team leader] by the country program advisors and country program team in West Africa to do what they call an output-to-purpose review [a summative evaluation]."

Evaluation Background

The evaluator described the evaluand as "a peer HIV education-based intervention implemented in secondary schools…. implemented by the state education authority in collaboration with a local NGO in West Africa … where the evaluation question was: 'Did the program meet its intended objectives?'" The evaluator stated, "In this instance, I had very little freedom to specify methods, approach, and so on. I was captive to the processes."

The evaluator described the data-gathering process as "consultations with children, with teachers, and with the community in a rather ad hoc way. Whom we talked to was dependent on the visits that we made rather than based on any planning in advance. We didn't have a framework suggesting, for example, that we visit five or ten schools and have a certain number of focus group discussions. We did not do that … I'm pretty sure, looking back that we were taken to high performing schools because what we saw was very positive."

The evaluator stated that the West African evaluation went smoothly, because the NGO had managed to document the processes and had done their own research to gather evidence of learning outcomes through KAP (Knowledge, Attitude, and Performance) surveys. The evaluation team "look[ed] at the evidence … ask[ed] some pointed questions ... [and] observ[ed]

classes being delivered in schools." The evaluator reported that the evaluation did not encounter "major difficulties … [was] confirmatory … [and] was a very positive evaluation."

Evaluator's Philosophical Assumptions

The evaluator stated that when conducting an evaluation, he often finds that "the methods, the consultations, and the outcomes are pre-cooked." He has found this type of approach to be "a very low-grade evaluation process because there is no real opportunity to test any of the hypotheses that might have been generated. You are less likely to get any major surprises with the one that's carefully controlled."

The evaluator indicated that when he conducts "a precooked evaluation," he worries "about bias because there was [is] bias built into it." He also said that he felt that his role "was not truly independent because I was a member of the funding agency." The evaluator stated that as an internal evaluator, "You work within the architecture and you work within the assumptions of the organization …. [On the other hand,] as an external evaluator, you have much more power and voice. You are able to pick up things that maybe people haven't seen …. You come in with a completely fresh perspective and with a critical mind … [and obtain a] different type of result." The evaluator stated that he felt that "in practice, no external evaluation is 100% external … [since] there are always perspectives that the agency will bring. It's usually a partnership."

The evaluator believed that "for evaluations to have integrity … there should be intellectual independence…. I think the concept of integrity is really important. You need to allow for the possibility that the project or program is not successful…. And that's what agencies do not like." The evaluator stressed the need for an "enabling environment for an evaluation … [that allows] evaluators to ask difficult questions." The evaluator stated, "… to set up the

enabling environment for evaluation … it's good to have some theory underpinning your evaluations—to give it rigor, but to also communicate that you are following a specific methodology."

The evaluator emphasized that "the organization [commissioning the evaluation] needs to provide me the opportunity to select the evaluation team." The evaluator recalled, "I have done single evaluations on my own, and it's extremely difficult to do…. You need other brains to bounce ideas off. You need to off-load some of the field work and so on." When he lacks "the prerequisite local knowledge," the evaluator reported that he relies "on the organization's [commissioning the evaluation] suggested local teams to help him with the evaluation."

The evaluator pointed out that "as a team leader, you need to identify who is going to be the principal investigator in the evaluation…. Otherwise, the team leader is someone who ends up writing the report or who has to deal with problems as they emerge rather than someone who is responsible for the intellectual groundwork as well as driving and guiding the process."

The evaluator said that "when you do an evaluation, you need to do a lot of preparation. You need to read the documents the agency has provided, and then you need to conduct a literature review." The evaluator also emphasized that "documentation needs to be given in advance … [and] needs to be well organized and managed by the organization that is commissioning the evaluation … [but instead] what they often do is dig [document files] out, then they send them to you in no particular order and in zip files." He made the recommendation that "the evaluator should have the space, with the organization, to specify how he or she would like the information packaged."

The evaluator highlighted that "evaluations depend on the quality of the project document—the logical framework." The evaluator advised, "in a good quality document, the

objectives are very clearly specified, the indicators are appropriate, the outputs are very clearly specified so that you can measure them. In a poorly designed document, the objectives are vague, unmeasureable [sic], unquantifiable, and not accessible or require too much effort to access them. Their indicators are inappropriate." The evaluator counseled, "I've seen both, and the second is extremely difficult."

When he is designing an evaluation, his goal is for the "evaluator and the key stakeholders to come together to actually decide on the evaluation approach … [because] the approach should depend on the type of evaluation that is asked for." The evaluator declared that "[the type of evaluation] should take some time to negotiate … because the organization that is implementing the program will have its own decision points … [such as] when to hold the evaluation [and] how much money has been budgeted." The evaluator indicated that "[the evaluation] is often under-budgeted … [and] there is very little scope to increase the budget. So that is an issue." The evaluator warned that evaluators "shouldn't simply bring in a template…because every [evaluand's] context is different."

The evaluator stated, "I think every project … has its own secret history which you need to grapple with. Every evaluator needs to be a detective and an anthropologist to try to understand how the project came into being and to understand its ecology.… You have to try to identify what drives that ecology…. You need to identify the trends within and who has power and who does not have power …. Often the evaluation just becomes a political exercise: 'Protect the project. Protect the agency. Protect the Ministry of Education.'"

The evaluator reported that he has found that "HIV health education in many countries is quite marginal. It depends on how much priority the Minister of Education is giving and wants to give to HIV in the school house …. The power you have behind you in your agency … depends

on a commitment of the government." The evaluator stated, "[The commitment to HIV health education is] changing over time. There are agencies that are interested, but I think many international development agencies are less committed."

The evaluator stated that "if you want to get the maximum effect after an evaluation, you need a process that engages the Ministry before anybody hits the ground. You need to get them thinking about it. You need to remember that they are busy people." The evaluator advised that "it is also important to meet key stakeholders early." The evaluator stated that engagement needs to "happen *immediately*. Otherwise, it will lead to very bad process. Governments don't like people to arrive and start work without having engaged with them."

The evaluator stated that for this evaluation, the evaluation research "was qualitative research. Not high-grade qualitative research at that. We were simply talking to people. The agency coordinated which sites we were to visit." The evaluator recognized that "it would have been good to have visited a range of different schools—good performing schools, some middle, and some poor." The evaluator recommended, "You need to set out your criteria for sampling the schools. That way [evaluators] could get some idea of the range of the implementation." The evaluator said that approach would have been much "more systematic [and] more scientific."

The evaluator was concerned about how the evaluation was conducted because "… apart from the site visits, there wasn't an opportunity to triangulate any of the data." The evaluator said, "I think it's important for evaluators to try to aim towards the scientific as far as possible. And it is difficult because of the imperatives that you are working under. Very few of the institutions that you work with are really geared up to doing this."

The evaluator narrated, "I have done evaluations where I have presented and left. [Then, due to time constraints,] I had to work about eighteen hours a day to get the report done. You

know, it can be *so compressed*." The evaluator advocated that it is important "to come back and to present the final version … because the report you submit when you leave and the report that's finalized can have some quite significant differences because you had time to reflect."

The evaluator concluded his story by saying, "You are always up against the issues of time, resources, and the extent to which you are able to control or to manage the process, especially if there are many actors involved. I find every evaluation builds up my understanding of how to do an evaluation. It's interesting—no matter how many times you do one, you learn something from every process. There is always a lesson to be learned."

Critical Incidents

Inability to provide input into the evaluation design. The evaluator stated that "the agency had set up the evaluation before I arrived. It did not give me the opportunity to provide any meaningful input in the design of the evaluation. The design was shared with me, of course, at a distance." He stated that "… the questions were already pre-cooked. Their methodology asked, 'Have the outputs been achieved on time, etc?'" The evaluator reported, "We had to work within the guidelines that were given to us. The guidelines are quite broad and generic. They weren't specific to HIV and school health."

According to the evaluator, "the evaluation was strongly geared to the logical framework…. So, in a sense what you are doing is checking to see whether the outputs have been realized, checking to see whether data have been obtained against the various indicators, and making some judgment as to whether the program's purpose is likely to be achieved by the end of the project."

Hurried evaluation. The evaluator was concerned that "only three months [were] allocated to conduct the evaluation …. So did I have a chance to conduct a literature review for

the evaluation in West Africa? *No, no, no*. It was much more hurried. It was much more focused."

NGO and culture. The evaluator stated that "the main challenge was trying to understand the contexts and trying to understand how the NGO was operating. It was a very short visit to the country. It was less than two weeks. There were a few visits to the field, which were all the same—incredibly illuminating. It was my first trip to that country. It was my first acquaintance with their culture."

Non-negotiable budget. The evaluator was disturbed that "in this evaluation there wasn't much room to negotiate the budget. There was nothing we could do. I think it is a very useful step in the process to check that you've got enough resources. You need to make sure there is enough money to pay for whatever extra stuff you might need."

Evaluation commissioned at end of project's life cycle. The evaluator indicated that "there really wasn't very much time to do the evaluation…. The end of the project was coming up, so it was close to the wire. It was two weeks or less than two weeks in country [to conduct the field research], and then I return to the home office overseas [to write-up the evaluation report]. I had to write the evaluation report along with the rest of the work that I was doing, so there was very little time to focus on it. So they got a very short report."

Evaluator's Decisions

The evaluator found that the evaluation did not have a sufficient budget. He found that the evaluation team did not have enough funds, so they paid for the evaluation expenses through the evaluation funding organization's budget. He stated that he "was part of the funding organization, so we just managed it with our resources." He also reported that due to time

constraints, he needed to "find ways … [to] bring in other people [and] … *again*, that required extra budgeting."

The evaluator generated a short evaluation report. He "never really did a formal evaluation presentation ….We submitted a report, which was then circulated …. [That was because a presentation] requires another visit, which adds to the time and budget of an evaluation." The evaluator "recommended that a more detailed evaluation should take place … because it required a fresh look at it all." The evaluator stated that "the [evaluation] approach changed because we brought in a high caliber evaluator [for the next evaluation] … who you don't just tell him [the evaluator] simply to write up something."

Case Six: Design and Development of the Formative Evaluation of a Country's HIV/AIDS Home-Based Healthcare Supervision Programs

Evaluator's Profile

The evaluator in this case six was a woman with more than 10 years of experience in conducting health and HIV evaluations. She was born in Sub-Saharan Africa and currently resides in North America. She is a consultant who works at an international aid agency's world headquarters. She works with an evaluation team that provides technical assistance to organizations working in the field. She holds a Master's in International Aid Development (MIAD) and a Master's in Public Health (MPH). In this scenario, the evaluator was a part of an evaluation team that was providing evaluation technical support to a team located in Sub-Saharan Africa. The evaluator stated that the team's "goal was to teach them how to conduct systematic, sound, rigorous evaluations." The evaluator emphasized that the team "tried to facilitate an evaluation that is conducted with a purpose in mind and to make sure that the evaluation findings were useful." The team receiving technical assistance wanted to implement

an evaluation to determine whether the program being evaluated met the program's objectives and was being implemented properly but was under significant time constraints.

Evaluation Background

The evaluation technical support team was providing assistance to a local partner at a university in southern Africa. The local team was interested in conducting a process evaluation of all of their country's HIV/AIDS home-based healthcare supervision programs. A component of the program involved providing HIV/AIDS health education. The program included meeting with clients who were infected with or affected by HIV/AIDS and providing health education on how to prevent infection or how to get medical care for any issues related to their HIV status. The programs being evaluated were funded by the international aid agency that employed the evaluation team providing technical assistance.

Evaluator's Philosophical Assumptions

The evaluator indicated that her "organization's mandate is to increase evaluation capacity for program evaluation …[but the team does not] "immediately accept requests for technical assistance." The technical support team "implements a screening process" that includes conducting a readiness assessment to "ensure that there is actually a project than can be evaluated" and to "make sure there is someone who is working on the evaluation that will be committed throughout the process." The team also asks questions that include, "What are you trying to do? Do you have a project that is ready for evaluation? What type of technical assistance are you requesting? What skills are you interested in strengthening? What is it that you really want to know [from the evaluation]?" The evaluator further stated that "Once the questions are answered, we give them examples of program evaluation skills they might need,

and then we describe what the technical support process is like." The evaluator reported that based on their assessment, the "evaluation was going to rely heavily on qualitative data."

The evaluator explained that the team's "approach in building evaluation capacity is based on adult learning principles. We selected this approach because the learner gets to apply what they are learning … without us telling them what to do. We provided them with feedback." The evaluator stated that "we have learned from our experience that this approach works best."

The evaluator stated that "based on the many years of previous evaluation technical support experience ... and working with people with different understandings of M&E (monitoring and evaluation)," the technical support team adapted an existing workbook that "addressed issues such as evaluation ethics and the use of the participatory approach. It [The workbook] also described all of the different [program evaluation] frameworks and steps on how to use the different frameworks."

The evaluator emphasized that the technical support team is "concerned about stakeholder use or ensuring that evaluations are feasible and utilized." She further explained, "We use Michael Q. Patton's Utilization Focused Evaluation principles." The focus on evaluation use influenced the team's decision to "focus [the evaluation] on the lack of supervision standards in supervision of home-based care … [because] we felt that the evaluation had the potential to inform home healthcare standards."

When the local evaluation team submitted an unrealistic data collection timeline, the technical support team "felt that they had to put the evaluation on hold." The evaluator explained, "As part of our responsibility—and I think part of the ethics as evaluators—we had to say in good faith, 'We cannot go forward.' So we had to say, 'We need to see how you [the local team] are going to train the data collectors.'" The evaluator reported that they "had that

authority as indirect funders … to tell our colleagues at headquarters to put a hold on the project, or at least not let them proceed with data collection, until we have sent in [feasible] data collection training materials. We were concerned about the quality of the data that would be collected."

As the training continued, the local team became less communicative. The evaluator was troubled that the PI (Primary Investigator) was the only one talking. So, the evaluator "became very concerned that other members of the team, who were also Co-PIs on the project, [and] were not as vocal as they used to be." She stressed, "I think it is important to seek input from those you don't hear from. I think experience has taught me this. It makes the team stronger."

The evaluator emphasized that "it was important to ask, 'What is motivating this rush in the timeline? What is motivating this dominance? What is motivating this need to speak on behalf of everybody, not allowing people to provide feedback, and being defensive?'" She stated that it was important to ask these questions "in a way that doesn't trigger a negative response."

The evaluator stated that when conducting an evaluation, "communication is challenging on every level because we are all different people." The evaluator explained, "I have worked with people from so many countries, so many different cultures, with people at the community level. It is something I enjoy doing—it is part of my personality. I believe that it is important to learn how to work with people from diverse cultures. Over time I have improved my own people skills." Then the evaluator stated, "When you are providing technical support, it is important to know that there is an organizational culture in every organization. Every community has its own hierarchy, and that culture determines when people can speak, who can speak, and when."

The evaluator discovered that the PI was concerned about not having anything to report to headquarters for the semi-annual review. The evaluator revealed that "as program managers,

we always think of numbers or quantitative results as a way to describe progress, but qualitative data, such as narratives that can explain what has happened, is also reporting." She stated that "this information helped to alleviate the PI's anxiety."

The evaluator shared with the PI a description of what it takes to get an evaluation approved. According to the evaluator, bottlenecks occur because of the many "IRB reviews of the [evaluation] protocols." The evaluator stated, "Our agency is science based — evidence based — [so] there are a lot of studies and research going on. We probably don't have enough people reviewing protocols. In the midst of that, the reviewers have other responsibilities. In light of the amount of research that takes place on a regular basis—and compounded with bureaucracy—there are bottlenecks."

Critical Incidents

Unsatisfactory protocol. The evaluator began her story by stating, "Our organization's Sub-Saharan headquarters requested our assistance in helping the local partner to strengthen their evaluation protocol so that they could implement an evaluation of high quality. Once we took a look at the protocol that the local partners had put together, we realized that it was of low quality." The evaluator stated that "as a team, we didn't assume that they [the local team] needed help for everything. Originally they said they only needed help in the analysis, but working with them, we realized they needed a lot more help. We had to start from scratch."

The evaluator said that in this instance, the local organization "sent us products that they developed during the training, and we provided feedback. At one point, the product that was being sent—the data collection timeline—was not at a satisfactory level. You need to have a good data collection protocol. It is especially important when you go into data collection. It can be a nightmare when you don't have data of good quality. It can really give you more gray hair

when it is time to do the analysis and report the findings." The evaluator said, "There wasn't enough time built into the timeline to train data collectors, pilot test the instruments, review results from the pilot test, and mak[e] adaptations as necessary before starting data collection. That timeline was so short that we were immediately concerned."

Lack of national guidelines. The evaluator and the team decided to conduct an evaluability assessment. The evaluator reported that "as part of our protocol, we conducted a literature review…. We found that national guidelines on the supervision of home-based healthcare did not exist. We had to ask ourselves, 'How can the quality of supervision of home-based healthcare be evaluated if there are no standards to compare it to? To what are they going to compare their analysis to see if there was impact?' Based on our findings, we decided to ask the evaluation questions, 'How is the supervision of home healthcare being implemented?' [and] 'Is supervision being conducted? And if so, how?'"

Logic model and misunderstood program. The evaluator highlighted that the team "decided to develop a logic model of the supervision component of the program. Yet, no one seemed to be on the same page of what that logic model should look like. It was clear that there wasn't a clear understanding of the program … That needs to be established before trying to determine if, where, and when one should evaluate a program. It also needs to be determined before you can decide which aspects of the program should be evaluated."

The evaluator also said, "It wasn't clear if the program's healthcare supervision was being implemented the same way across all of the programs. That would make it *impossible* to answer the proposed evaluation questions. It was important that local partners ask themselves, 'Is the program implemented consistently?' and 'How is it being implemented?' These questions needed to be answered before they could determine if the programs were of good quality."

The evaluation technical support team decided to use a "workbook that was divided into four phases. One of the phases included readiness assessment—also called an evaluability assessment—to see if the program is at a stage where it can be evaluated. One of the steps assisted in the team to identify the evaluation stakeholders and to determine why it is important to define them. We provided them with tools, or a matrix, that they could use to help them identify the stakeholders' names and their organizations, the roles that they played, and their contact information."

Non-communicative Co-PIs and a domineering PI . The evaluator stated that the team "had established a system of having weekly calls, during which we provided feedback, discussed next steps, or provided follow-up as necessary. It was during one of those calls that we noticed that the PI was the only person talking. When people are quiet, I always worry. We noticed that the tone had changed…. The PI [Primary Investigator] was speaking on behalf of everybody."

She revealed that she found "out that the PI was concerned that she wouldn't have a report for the semi-annual review…. Program directors can easily feel that [way] when it is time for the semi-annual or annual report [and] there is nothing to report. They are pretty much on hold while the IRB reviews your protocol. It came up—after I asked some exploratory questions that were carefully worded—that the PI was feeling a lot of pressure because the program didn't have any results to report."

The evaluator said, "You know, this provided me an opportunity to share with them [the PI and the rest of the team] how long it took for us to go through an [previous] evaluation through our organization and how we lost staff during the process because it was such a long process…. As a result, we didn't have any results to report for our review. It put her at ease." The evaluator said, "I told her, 'You are not alone. It is not a reflection on you. There are other

dynamics that happen.'" The evaluator stated, "I know that people want to save face, so I said, 'I understand that you don't want to seem as if you don't know what you are doing.'" The evaluator stated that "we talked about creative ways to report results. It helped to let her know that different ways can be chosen to describe what has been done and still fulfill the reporting requirements."

The evaluator said that the evaluation technical support team "provided the team with training on leadership, management, and their roles and responsibilities as members of a team. Without a doubt, after we overcame the power struggle, everybody felt empowered. Everybody mentioned how life-changing this process was for them, and they used the word 'empowerment' as a way to explain … what they got out of the experience. That is why I am probably using it."

Bureaucracy and non-payment. The evaluation technical support team was very concerned when then "learned … that the institution that was funding the program had not paid [the local team's] salaries, due to their own [internal] bureaucratic issues. [The local team] worked for 18 months without getting a salary. That really affected the morale of the team. The amazing thing is that they brought it up, but they didn't dwell on it very much. They kept on doing their work. It was one of the things that came out when we were trying to find out what was really going on with this team. That is when we [the evaluation technical support team] found out about the payment issues. We pledged to do the best that we could to move that along, especially if there were hold-ups coming from our organizations."

"Even though they weren't getting paid, the local team continued working and sent us their PowerPoint presentation with 131 slides. They continued to send more information and continued to participate on our weekly conference calls. In spite of not getting paid, they were

still sending us products—and this is the whole team. Everyone was very active. If we got on the phone, everyone had something to say. That team has consistently worked with us. The sad thing is that as of today, due to in-country administration and bureaucratic issues that are impeding release of additional funds from the U.S. government, the project is on hold."

Evaluator's Decisions

The evaluator stated that to provide assistance, the team decided to conduct an evaluability assessment, provide technical support based on adult learning principles, and develop a workbook and tools. To address the lack of consensus regarding the program, the team assisted with the development of a logic model and also with the development of evaluation questions that would aid the development of home care supervision standards. The evaluation team delivered leadership and management training to address the issues regarding local team member participation and empowerment. Based on their work, the local team was at a point where data collection could occur. The evaluators also provided opportunities for the local team to express their concerns, such as the lack of payment.

Case Seven: Formative Evaluation of the Health Education Component of a Surgical Intervention That Reduces HIV Transmission

Evaluator's Profile

The evaluator in case seven was a male with more than ten years of experience conducting evaluations of HIV/AIDS health programs. He lives in the Middle East and is a senior international consultant working with an international health organization. He has extensive experience in research on HIV prevention and evaluation capacity building in Africa. His research also focuses on public health, health literacy and leadership, and HIV/AIDS

curricula. He has been involved in policy development, operational research, and M&E (monitoring and evaluation) and training of local hospital teams in medical male circumcision.

He is a lecturer at a school of public health and community medicine in the Middle East and the United States. He is also traveling faculty for an international pre-med program. He has a PhD in Public Health and Science Education, an MPH (Master's in Public Health), and a Bachelor's degree in Political Science and Sociology. He has been honored for his research on HIV/AIDS education and prevention.

The evaluator stated that he was a consultant hired by the intervention's funding agency to conduct an external evaluation of an intervention that utilizes male circumcision to reduce HIV transmission from infected females to their heterosexual male partners. His main responsibility was the development of the tools and the protocol. The goal was to develop a protocol and instruments that were sensitive to the context in which the evaluation was being conducted.

Evaluation Background

The evaluator stated that "controlled clinical trials conducted in three Sub-Saharan African countries have provided evidence that the removal of the foreskin reduces the transmission of the HIV virus from infected females to uninfected males by 65%. Based on the recommendations of the WHO, by UNICEF, by UNAIDS, and NIH, Sub-Saharan governments have adopted policies that include male circumcision as a means of reducing HIV/AIDS transmission, combined with behavioral change and education strategies. Male circumcision has moved from policy to legislation to practice as a part of Sub-Saharan Africa's HIV prevention strategy. Clinics are being opened, and thirty to fifty men a day, in their late teens to thirty-five years old, are voluntarily coming to be circumcised."

In 2008, Jewish and Muslim practitioners started to train local medical teams in a Sub-Saharan African country on how to perform high volume circumcisions that assist in the reduction of HIV transmission from infected women to their heterosexual male partners. The intervention's funding agency required that a monitoring and evaluation component be included in the core program, or else the program would not be approved. An evaluation of that program was concluded in 2009.

Several results were found from the evaluation in 2009. First, it was found that after a very sophisticated, hands-on, five-day course, a local cadre of two or three doctors, nine to twelve nurses, and two administrative staff members could be fully trained to be able to circumcise thirty clients per day. The trainers flew in from overseas, landed in-country, provided the training, and, as a result, the local teams were performing surgery on at least 25 to 30 clients a day. These numbers have been maintained up to now, and with more experience, they could meet the needs of 50 clients per day. Within nine months, seven such courses have been conducted in the community that we plan to evaluate in 2011, and three courses have been taught in a neighboring country, which were evaluated in 2009. As a result, there have been ten courses. So in total, what this program has done is to move the medical teams from zero circumcisions to possibly 50 circumcisions per day. If you estimate two hundred working days per year, we are talking about at a minimum of ten thousand men per team being circumcised. With the ten teams that have been trained so far, we are talking about one hundred thousand men circumcised by graduates from our ten training courses.

"In 2009, we began to implement the second intervention in another African country, located in Sub-Saharan Africa, with a population of at least 10 million people." The evaluator indicated that "the educational component of the intervention began from the moment the client

entered the circumcision clinic, continued throughout the operation (about eighteen to twenty minutes), through to recovery (about thirty to forty minutes post operation where the nurses monitor the client), and with take-home materials. So that meant that the educational component was split into four parts."

The evaluator said that "it was great that the local leaders, the government, and health department had created an enabling environment, but we also needed to do so some outreach and educate the men about the procedure." The evaluator stated that "at the entrance to the clinic, posters and leaflets were available for men interested in the surgery. Since the men were awake during surgery, which was conducted under local anesthesia, the clients could read while the surgery was performed. They were handed a well designed brochure about circumcision and how to take care of the incision. (The literacy rates are very high in this community, so literacy is not a problem.) As the men were recovering from surgery, we complemented the intervention with another level of education. As they were being looked after by the nurses, the men read through a cartoon booklet with an emphasis on HIV prevention. When they walked out of the clinic, they were provided with some pain killers and were handed educational materials to read at home."

Evaluator's Philosophical Assumptions

The evaluator stated that when the team was designing the educational materials that were to be used in the intervention, they asked themselves, "'How can men, after the procedure, walk out of a clinic better educated about the healing process? How can we assist them in appreciating that the procedure does not provide 100% prevention? How can we emphasize that it only dramatically reduces the chances of them contracting the virus?'" The evaluator stated that "it was a complex educational intervention that was part of the HIV intervention."

As an evaluator, he felt that when evaluators conduct an evaluation, they "bring some of the basic theoretical frameworks that are common practice." The evaluator stated that "in this particular intervention ... I felt that the cross-cultural elements needed to be addressed at three levels. Level number one is that all the clients were very homogeneous. They were men from the same tribe, spoke the same language, and they had the same socialization. Still, because some of the men came from urban areas and some came from more rural areas, we needed to look into subcultures within the tribe's main cultural framework. Number two was the interaction between the medical practitioners, the hospital/outpatient community clinic, and clients. There was an interaction between the healthy people and medical people—those who serve them and test them, research them or educate them, or perform surgery on them. And the third cultural issue was that the doctors were coming from a very foreign country, very well developed country, usually were whites, in most cases on very high economical strata, who were coming to work and to teach often black, less well-to-do economically, group of very eager medical professionals in Africa. So the clients' and medical personnel's cross-cultural, cross-religious, cross-racial, and socio-economic status needed to be addressed in the research instruments and the protocol."

The evaluator stated that "my contribution was significant in that regard because of my experience as a former UN employee, a past UNESCO staff member, as an evaluator, a researcher, and as a public health practitioner in the field." The evaluator stated that in this evaluation, "we used a mixed method approach because the intervention was so novel. In addition, a closed questionnaire gives you one kind of information, and open interview gives you other information. By using both methods it gives strength to the evaluation."

Critical Incident

Geographical barriers. The evaluator stated that "throughout the evaluation process we communicated through email and met via Internet. It made life easier to use the technology at our disposal. It was a way to bridge over geographical gaps."

Cross-cultural and language barriers. The evaluator stated that, based on the lessons learned from an evaluation of a similar program conducted in 2009, the team refined the evaluation tools for this intervention's evaluation. The evaluator said, "We knew that the context did not transfer exactly from one evaluation site to another. We were aware that the messages needed to address the men's different ages, social class, and literacy rate, as well as their professional or laymen background." The evaluator further explained, "For example, we initially developed a draft of the cartoons that were going to be used to educate the circumcised men about the healing process and the significant benefit of male circumcision. They were originally drawn with characters which were very western because the cartoonist was not an African person but a western person."

Language and culture. The evaluator stated that the "[HIV/AIDS health education] text was written in English. In a way, that was okay because English is a very common language in the two countries where the intervention was implemented. [English is] better understood than some of the local dialects. So there was no problem with the use of the English language in the materials, but there was a problem with how the words were used. This was because in the country's context, English words are sometimes used differently than in a western country."

The evaluator explained that "the male nurses were sometimes explaining the materials that were originally in English into their local language. We were curious as to why they needed to do that because generally the population spoke and understood English. The nurses explained

to us that some words do not exist in their local language. They also explained that some words have a more powerful meaning in their local language, so they explained the concepts written in English in their language."

Political support. The evaluator stated that in 2009, the practice of male circumcision was supported by the country's leaders and the department of health. As a result, there was an enabling environment that created a huge demand for the service of male circumcision.

Circumcision's public health role. The evaluator stated that "while there is a stigma associated with HIV/AIDS, I want to highlight that without our needing to intervene; stigmatization against religion was not an issue in the field. This was because recipients of the training and the clients disassociated circumcision from its religious role to a public health role. So the religious aspects associated with the surgical intervention to prevent the transmission of HIV/AIDS seemed to have worked themselves out."

Logistics and methodology. The evaluator said that "what really helped is that we had a lot of time, in terms of logistics. We were not pressed with time issues…. Our situation wasn't such that we had only one hour and we had to decide 'Do we do a questionnaire or an interview?' We had ample time to execute different methodologies for the evaluation, so that helped us with strengthening the evaluation."

Successful collaboration with stakeholders. The evaluator stated, "We worked very collaboratively. We were a multicultural and diverse group of professionals that assisted in governing the project. Everyone in the team brought in their own experience and expertise to better massage the cultural challenges. It was beautiful to see how quickly we were on the same page. There weren't any dramatic issues. No, everybody was contributing to the development of the instruments and to incorporate the specific cross-cultural issues. When it was appropriate,

we worked together to write the survey and interview questions. We further collaborated as we determined when to ask the young men specific questions. We worked as a team. As I previously indicated, the evaluation is not complete, but I wanted to share some of the parameters and first results coming from operational research on male circumcision as a surgical intervention to prevent HIV infection."

Evaluator's Decisions

The evaluators decided to use technology to overcome the geographical barriers that occurred because the evaluators were located in different parts of the world. To address cultural issues, the educational and promotional materials such as brochures, posters, drawings, and even pictures were pretested to make sure that they were culturally acceptable and effective. The evaluator stated, "We developed a questionnaire to determine the messages' acceptability." The evaluator reported that the evaluation team "developed questions to determine whether the participants understood the messages, text, and the cartoons. One of the things that I did was to defend the use of a local team of people from Africa so that they could recommend and provide changes to the drawings of the body in the educational material so that the person in the drawing would look African."

To address language barriers, the evaluation team "worked with the in-country group to incorporate some local language to enhance the educational messages. I would say that 90% of the text was in English and 10% was in the local language. Through interviews, we provided the responders the opportunity to share their thoughts, feelings, and assessment of the effectiveness of the materials that were presented to them."

Summary Overview of the Seven Cases

In this section, tables are utilized to depict the scope, variety, range, and depth of the data present in the seven cases in which the expert evaluators evaluated HIV/AIDS health education programs implemented in Sub-Saharan Africa. The tables are designed to provide a summary overview of the seven cases to depict the participants' demographic information, types of evaluations conducted, evaluation funding agencies, and a high-level summary of each of the cases.

The research participants' profile provided in Table 4.1 illustrated the scope, balance, and variety of demographic data present in the seven cases. The data from this table were gathered from the interviews, the co-constructed member-checked narratives, and timelines, and from the résumés and biographies that the evaluators provided.

The first column in the table listed the evaluator's demographic information: gender, evaluation experience, academic training, and residence. Columns two through eight listed the participant's case number. For example, the row labeled *Gender* provided a description of the number of female (3) and male (4) expert evaluators depicted in the seven cases. Reading a case's column provided the profile of a research participant. For example, in Case One, the evaluator was a man who had ten years of evaluation experience, obtained a PhD and an MPH, and resided in Sub-Saharan Africa. This table indicated that evaluators had at least 10 years of evaluation experience and graduate degrees, and three resided in Sub-Saharan Africa, two in North America, one in Asia, and one in the Middle East.

Table 4.1

Research Participants' Profiles

	Research Participant						
Characteristics	1	2	3	4	5	6	7
Gender	Male	Female	Male	Female	Male	Female	Male
Experience	10	19	12	12	30	10	10
Academic Training*	PhD MPH	PhD MPA	M.H.S.	PhD MPH	PGCE MA	MIAD MPH	PhD MPH
Residence	Sub-Saharan Africa	Sub-Saharan Africa	Sub-Saharan Africa	North America	Asia	North America	Middle East

*PhD = Doctor of Philosophy, M.H.S. = Master of Health Science, PGCE = Postgraduate Certificate of Education, MIAD = Master in International Aid Development, MPH = Master of Public Health, MPA = Master of Public Administration, MA = Master of Arts

Table 4.2 was utilized to illustrate the scope, balance, and variety of the types of evaluations and the different evaluation criteria represented in the seven cases. The data from this table were gathered from the co-constructed narratives. The first column in the table indicated whether an evaluation was a formative or summative evaluation, with a subcategory listing the evaluation criteria used to evaluate the cases. Columns two through eight listed the case numbers. The row labeled *Formative Evaluation* indicated that four of the seven evaluation cases were formative evaluations. The row labeled *Summative Evaluation* indicated that three of the seven evaluation cases were summative evaluations. Reading down a case's column provided a profile of the individual case type and the evaluation criteria. For example, in Case 1, the evaluation was a formative evaluation that utilized the following criteria: effectiveness, efficiency, and fulfillment of objectives.

Table 4.2

Evaluation Type and Evaluation Criteria

Evaluation Description / Criteria	Case Number 1	2	3	4	5	6	7
Formative Evaluation	x			x		x	x
Developmental				x			
Effectiveness	x					x	x
Efficiency	x					x	x
Evaluability						x	
Fulfillment of Objectives	x					x	x
Relevance				x			
Summative Evaluation		x	x		x		
Impact		x	x		x		
Sustainability		x					

A summary table indicating whether an evaluation was an external or an internal evaluation was provided in Table 4.3. The table indicated that the seven cases depicted a mixture of internal and external evaluations. The data from this table were gathered from the two interviews and from the co-constructed, member checked narratives and timelines. The first column header in the table labeled *Evaluation Type* was divided into two categories: External and Internal. The second through the eighth row headers list the case numbers. The row listed under the “External” category indicates that five cases were external evaluations, and the row data under the “Internal” category indicates that two cases were internal evaluations. Interpreting the table by reading the columns indicates whether a case is an external or internal evaluation.

Table 4.3

Evaluation Type: External and Internal Evaluation

Evaluation Type	Case Number						
	1	2	3	4	5	6	7
External	x	x			x	x	x
Internal			x	x			

A summary of the evaluation design and data-gathering instruments utilized was provided in Table 4.4. The table illustrated that the seven cases depicted a variety of evaluation designs and data-gathering instruments. The data from this table were gathered from the two interviews and also from the co-constructed narratives and timelines. The row data listed under Evaluation Design indicate that of the seven cases, two utilized a causal/quasi-experimental design, and all included a descriptive design and utilized an exploratory design.

The first column header in the table was categorized into two subcategories: Evaluation Design and Data-Gathering Instruments. Column headers two through eight list the case numbers. The data listed under Evaluation Design indicate that of the seven cases, two utilized a causal/quasi-experimental design, all included a descriptive design, and one utilized an exploratory design. By interpreting the Data-Gathering Instruments sub-category, the types of data-gathering instruments that were used in the seven cases indicate that four cases relied on surveys, all seven used interviews, and one utilized surveys. The data in the table also indicate that three cases used two designs, and five cases used two data-gathering instruments. Individual case information was also provided in the table. For example, the column labeled Case 1 indicates that the evaluation design was descriptive and that the data-gathering tool utilized was an interview.

Table 4.4

Evaluation Designs and Data-Gathering Instruments

Design and Data-Gathering Instruments	Case Number						
	1	2	3	4	5	6	7
Evaluation Design							
Causal/Quasi-Experimental			x				x
Descriptive	x	x	x	x	x	x	x
Exploratory		x					
Data-Gathering Instruments							
Survey		x	x			x	x
Interviews	x	x	x	x	x	x	x
Observation				x			

Table 4.5 illustrated the variety of development agencies depicted in the seven case descriptions. The list of international aid agencies was derived from the two interviews conducted for each of the seven cases. The specific international aid agency sponsoring the shared evaluations was not provided in order to protect the research participants' confidentiality, as the majority of the cases depicted high-profile evaluations. Each of the evaluations was funded by at least two international agencies.

Table 4.5

Agencies that Commissioned Evaluations

International Aid Agencies
DFID - Department for International Development
PEPHAR - The U. S. President's Emergency Plan for AIDS Relief
USAID - United States Agency for International Development
Global Fund - Global Fund to Fight AIDS, Tuberculosis, and Malaria
CDC - Centers for Disease Control and Prevention

International Aid Agencies
UNAIDS - Joint United Nations Programme on HIV/AIDS
WHO - World Health Organization
UNICEF - United Nations Children's Fund
UNDP - United Nations Development Program
UNESCO - United Nations Educational, Scientific, and Cultural Organization

Table 4.6 provides a summarized description for each case. The description showed the variety of evaluations depicted in each case. Column one, labeled *Case Number and Title*, lists the case numbers and the titles of the cases, respectively. Column two, labeled *Narrative Summary*, provides the summary of each case. The table rows identify the case's number and title and illustrate the summary for a particular case.

Table 4.6

Summaries: Seven Case Narratives

Case Number and Title	Narrative Summary
Case 1: Formative Evaluation of an HIV Health Department and an Interview of an Evaluation Department Official	We were commissioned to evaluate a country's health department programs. As the team leader, we were assigned to interview health education program managers/officials. We needed to interview the senior government official in charge of the HIV health program with an HIV health education component. When trying to interview a program management official, we encountered the manager's resistance to being interviewed.
Case 2: Summative Impact Evaluation of an HIV/AIDS Health Education Program's Consortium Model	Rather than individually funding HIV/AIDS health education organizations, a donor had the idea that a consortium would provide better results. The donor contacted me to do an impact HIV/AIDS prevention program evaluation with an HIV/AIDS health education component implemented in southern Africa. They said they wanted an impact evaluation that investigated the consortium model as a way to fund HIV/AIDS health education programs. The donor refused to explain their understanding of an impact evaluation, their anticipated use of the evaluation, and who would be using the evaluation. I needed this information to

	understand their perspective. They just did not want to talk to me.
Case 3: Summative Impact Evaluation of Micro-Credit and an HIV/AIDS Health Education Intervention	An NGO was providing micro-credit to young girls to reduce their involvement in risky sexual behaviors that could lead to exposure to HIV. We convinced the NGO to allow us to implement HIV/AIDS health education intervention through their program. We decided to conduct an impact evaluation of the program. The NGO had already made promises as to who would participate in their program, so our hope of conducting a randomized control study was not possible. Due to our inexperience (in conducting this type of evaluation), we asked the NGO to assist us. They indicated that they had the necessary research skills, which would assist us in the evaluation. The data collected indicated that the NGO was not equipped to assist in the data collection process. We had submitted our proposal to our organization's headquarters for approval. We thought we had received approval to conduct the study, but due to a bureaucratic mix-up and our lack of experience in completing the organization's approval process, the qualitative portion of our evaluation was halted.
Case 4: Ex-Ante Evaluation That Was Aimed at Designing a Transactional Sex and Women's HIV/AIDS Vulnerability Evaluation	We are conducting a formative evaluation in two countries. We were doing the formative assessment first so that we can design the intervention's educational component. The intervention's goal was to understand how heterosexual transactional sexual relationships were initiated or sustained and to use that information to design a HIV/AIDS health education intervention. We are not studying commercial sex workers. To understand the evaluation context, we decided to conduct unstructured observations. We met with key stakeholders as a part of our engagement and the community entry process. In one of the countries, civil unrest broke out, so we had to pull out. In the other country, the Ministry decided that they would not allow any human subjects research. As a result, we placed the evaluation on hold due to what was occurring in the two countries.
Case 5: Summative Evaluation of a Secondary School HIV Health Education Intervention	As an internal HIV/AIDS education evaluation team leader for a government-based development funding agency, I was brought in by the country's program advisors and country program team to conduct an output-to-purpose review. The program I was evaluating was a peer HIV health education intervention implemented in secondary schools. It was implemented by the state education authority in collaboration with a local NGO. The agency had designed the evaluation beforehand and did not allow me input into the evaluation design. It was a rushed evaluation. Since we had never worked in that country, it was difficult to understand the evaluation context. The budget for the evaluation was so small that we barely covered our expenses. I had to rush to write the report, and we did not present the findings.

Case 6: Design and Development of the Formative Evaluation of a Country's HIV/AIDS Home-based Healthcare Supervision Programs	As a part of an evaluation technical support team, we were supporting a local partner at a university in southern Africa to evaluate all of the HIV/AIDS home-based healthcare supervision programs in their country funded by our organization. Aspects of the program included the supervision of HIV/AIDS health education programs where the program implementers met with clients who were either infected or affected by HIV/AIDS. Our organization's African headquarters requested our assistance in strengthening the local partner's evaluation protocol so that they could implement a high-quality evaluation. Once we looked at the protocol that the local partners had put together, we realized that it was of a low quality. We also determined that the team did not have a unified understanding of the program's theory of change. We determined that we needed to modify the existing evaluation training material to meet their needs. We assisted them as they conducted an evaluability assessment of the program being evaluated. We then realized that they also needed help with data collection.
Case 7: Formative Evaluation of the Health Education Component of a Surgical Intervention That Reduces HIV Transmission	My team and I are evaluating an intervention that introduces surgery—male circumcision—to reduce HIV transmission from infected women to their heterosexual male partners. Fourteen countries in SSA are implementing male circumcision services that are geared towards thousand and tens of thousands of men who want to be circumcised. We are conducting a formative evaluation in one of those countries. Clinics are being opened, and thirty to fifty men a day, in their late teens to thirty-five-year olds, are voluntarily being circumcised. We pre-tested the HIV/AIDS educational material used to educate the men who registered for circumcision, had the circumcision, and were recuperating from the surgery. We also pre-tested the evaluation instruments for their effectiveness.

Chapter Summary

The first section of Chapter Four presented case descriptions for the seven evaluation cases. The cases focused on seven of the eight participants' stories based on their evaluation of an HIV/AIDS health education program that was implemented in Sub-Saharan Africa. The case description for the eighth case was not provided, as it did not bring further insight to the study. The seven case overviews were generated from the expert evaluators' member-checked, co-constructed narratives. Each of the case descriptions utilized the following format: expert

evaluator's profile, evaluation background, evaluator's philosophical assumptions, critical incident(s), and evaluator's decisions. The second section in Chapter Four contains a narrative description of the summary overview tables that described the participants' demographic information and types of evaluations conducted, as well as the evaluation designs and data-gathering instruments utilized. The diversity of programs evaluated, evaluators, and milieu represent a balanced representation of real world evaluation cases. The second section also provides a list of the international aid agencies that funded the evaluations and a high-level summary of each of the cases. Chapter Five depicts the interview process, how the narratives and timelines were generated and co-constructed, and the data analysis procedure used in this study.

CHAPTER 5

DATA COLLECTION, DATA REPRESENTATION, AND DATA ANALYSIS

The purpose of this chapter is to present a detailed insight into the protocol utilized to generate the narrative descriptions and graphical representations of the cases, and for data analyses. The chapter explains how the cases were developed. It also describes the procedures used for the following: data collection, narrative co-construction and member checking, decision making timeline co-construction and member checking, and data analysis. The three phases of Cognitive Task Analysis, the Critical Decision Method, Narrative Analysis, and Phenomenographic Methods guided the (1) data collection, (2) co-construction and member checking of the decision making timelines and narratives, and (3) data analysis. The research conducted by Klein (1993) and Kundin (2008, 2010) offered assistance in the current analysis of the research participants' decision making processes.

Section one herein explains the chapter's purpose and structure. Section two provides a description of the preliminary activities that occurred prior to data collection. Section three provides a description of the knowledge elicitation, knowledge construction, and the data analyses process. In addition, tables, diagrams, and quotes from narratives have been used to depict the processes used in my study. Chapter six provides the results based on the analysis of all of the seven cases in this collective case study.

Chapter Purpose and Structure

In this chapter, a description of the protocol I used in my study is provided to explain how my research was conducted. In addition, I provided this chapter because describing my

research protocol "is a major way of increasing the *reliability* of case study research" (Yin, 2003, p. 67). After I reviewed the seven cases, Case One was selected as the illustration case in this chapter for the following reasons: it was not as complex as the other cases; because the case focused on two critical incidents; and for the most part, Case One followed the order in which the questions were listed in the case study protocol. A review of the cases indicated that of the seven cases, Case One, in spite of containing a few variations from the original protocol questions, most closely followed the protocol, wording of the interview questions, and sequence. However, one major deviation from the protocol did occur because of a poor telephone connection. To comply with IRB requirements, quotes from the co-constructed member-checked narratives were used instead of direct interview quotes.

According to Yin (2003), a case study protocol contains "the instruments as well as the procedures and general rules" that focus on gathering data from a "single-case study" or a "single respondent" (p. 67). Yin (2003) indicated that a case study protocol should include the following: a description of the case research goals, the procedures used when conducting the study, and the questions used in the study. Further, Yin (2003) also stated that the "nature of the interview is more open-ended" in case studies (p. 72) in comparison to that in a survey, and consequently, progressive focusing can occur (Parlett & Hamilton, 1976; Stake, 1995). It should be noted that I did not ask the interview questions exactly as listed in the protocol, as progressive focusing occurred during the interviews. I found that the majority of the research participants were very skilled in sharing their narratives, and often some of the interview questions were answered before they were asked. This observation was consistent with Yin's (2003) statement, "In a case study, you must therefore learn to integrate real-world events with the needs of the data collection plan" (p. 72).

In such instances, I used the interview guide as a "checklist during the interview to ensure that all of the relevant topics are covered" (Patton, 1987, p. 111). I have used Case One to provide an example of the interview questions that were "adapted both in wording and sequence of questions to specific respondents in context of the actual interview" (1987, p. 111). It should be noted that during the interviews, one of the goals was to "[(understand how evaluators' made naturalistic decisions)], learn their [evaluators'] terminology and judgments, and capture the complexities of their [evaluators'] individual perceptions and experiences" (1987, p. 115).

Similarly, at times it was observed that the CTA's three phases overlapped. For example, data analysis was initiated in the first interview when I first began creating the drafts of the decision making timeline and the narratives. As stated by Patton (1987), "There is typically not a precise point at which data collection ends and analysis begins. Nor, in practice, are analysis and interpretation neatly separated" (p. 144).

Preliminary Research Activities

In this section, the preliminary research activities I conducted prior to initiating the research are described. The preliminary activities are provided to increase the reliability of the findings in my study and to indicate that there were several institutional reviews of the study's protocol. Two IRB applications were submitted to the author's research institution. In 2010, I submitted an IRB for a pilot study, which was approved in March 2010. The pilot study focused on determining the factors that influence expert evaluators' naturalistic decision making. I used the pilot study as an opportunity to test the research protocol. The findings (See Appendix A) were presented and reviewed by my dissertation committee. After fulfilling the requirements for the comprehensive exam and prospectus, I submitted a modified IRB application (See Appendix

B) in January 2011, so that I could conduct my research on how the expert evaluators' made naturalistic decisions. The modified IRB application was also approved.

The IRB process required that I provide a protocol that "covered the background information about the project, the substantive issues being investigated, and the relevant readings about the issues" (Yin, 2003, p. 69). As indicated by Yin (2003), the protocol served as an instrument that helped me maintain a focus "on the subject of the case study… [and] anticipate several problems, including [identifying] the way that the case study reports are to be completed" (p. 69). Both studies — the pilot study and the modified IRB — used a consent form that was approved by the IRB office at my research institution (See Appendix C).

CTA's Three Phases, the Critical Decision Method, and Narrative Analysis

Data collection, co-construction, validation through member checking, and analysis procedures were guided by the three phases of the Cognitive Task Analysis (CTA) and the Critical Decision Method (CDM) (see Figure 3.1). The research design I used in my study is depicted in Figure 3.2. A matrix is provided in Chapter Three, which lists the interview questions, and Sweeps One through Six (see Table 3.2). The majority of the data analysis was conducted during Sweep Seven. During data analysis, I utilized the narrative analysis procedure recommended by Creswell (2007) (See Figure 3.3).

This section explains the three phases of Cognitive Task Analysis: Knowledge Elicitation, Knowledge Representation, and Data Analysis. Each of the phases had unique goals, questions, results, and corresponding critical decision method sweeps. The interviews utilized the Critical Decision Method in Sweeps One through Six. In addition, the purpose of each sweep, the sweep's interview questions, and the research participant's responses are provided.

During the first interview, Sweep One involved introducing my research study and explaining the IRB-approved consent form. My study followed IRB requirements, such as explaining the consent form and obtaining the research participant's signed consent. I used the consent form to introduce and explain the study to the research participants. For example, in Case One, I explained the goal of the study — to understand how the expert evaluators make naturalistic decisions when evaluating HIV/AIDS health education programs implemented in Sub-Saharan Africa. In addition, I shared the overall research protocol with the research participant. Before the first interview, I emphasized that I was creating a rough draft of a decision making timeline to help determine the chronology of the narrative's events while the research participant was sharing his or her story. I also explained that a decision making timeline as well as a summary of the narrative would be provided at the end of the interview; that time would be allocated for clarification or correction of the narrative and decision making timeline. In addition, an electronic version of the timeline and narrative would also be provided before the second interview so that the research participant could review these documents beforehand. When this process was not interrupted by technical problems, I found it to be an efficient plan for generating the rough drafts of the narratives and decision making timelines as well as for preparing the research participant for the second interview.

During the second interview, I reviewed and modified the decision making timeline based on the research participant's feedback. The feedback obtained from the decision making timelines were incorporated into the narratives. In addition, a review of the narratives was also conducted. The research participants' feedback was incorporated into the narratives. This process of reviewing the narratives and decision making timelines as well as the co-construction and validation of the timelines continued until the research participant and I agreed that the

narrative was accurate and complete. This process proved to be quite effective in finalizing the decision making timelines and the narratives.

Knowledge Elicitation

During the interview process, knowledge elicitation provided me the opportunity to obtain information that focused on one or more critical incidents present in the narratives shared by the research participants. The research participants shared stories by recalling an experience that occurred while they were evaluating HIV/AIDS health education programs implemented in Sub-Saharan Africa. The knowledge elicitation process spanned interviews one and two. I used Sweeps One through Three of the critical decision methods for knowledge elicitation.

Sweep One: Interview One. Sweep One, focused on (1) explaining the purpose of the study, (2) explaining the consent form, and (3) gathering demographic information about the research participants. The Sweep One's purpose and interview questions are listed in Table 5.1.

Table 5.1

Sweep One: Purpose and Interview Questions

Purpose	First Interview Questions
Gather Demographic Information	How many years of experience do you have in evaluating HIV/AIDS health education programs?
	How many of these programs were situated in Sub-Saharan Africa?
	How many of these programs were funded by international aid agencies?

The first question I asked was: "How many years of experience do you have in evaluating HIV/AIDS Health Education Programs?" Case One's research participant stated that he had over ten years of evaluation experience. The second question I asked was: "How many of these programs were situated in Sub-Saharan Africa?" The research participant indicated that the

majority of the evaluations he conducted were in Sub-Saharan Africa. A review of his résumé indicated that the research participant had conducted over 12 HIV/AIDS program evaluations from 1999 through 2011. Some of the evaluations were related to his ongoing research projects. The third question I asked was: "How many of these programs were funded by international aid agencies?" The research participant did not give a specific number for the evaluations he had conducted that were funded by international aid agencies. However, a review of his résumé/curriculum vitae indicated that the majority of the evaluations he conducted were funded by international aid agencies and in collaboration with an academic institution located in Sub-Saharan Africa.

Sweep Two: Interview One. Sweep Two had two purposes. The first purpose was to answer the research question: What types of critical incidents do evaluators encounter during evaluations of HIV/AIDS health education programs? The other purpose was to identify a critical incident that the research participant, an expert evaluator, had encountered. Sweep Two's purpose and the corresponding interview questions are listed in Table 5.2.

Table 5.2

Sweep Two: Purpose and Interview Questions

Purpose	First Interview Questions
Critical Incident Identification	Can you think of a time when your skills as an HIV/AIDS program evaluator were really challenged?
	Can you share with me a time when your skills as an evaluator of HIV/AIDS programs really made a difference?
	Can you share with me an instance when an HIV/AIDS program evaluation would have gone differently if you weren't there?

Purpose	First Interview Questions
	Can you tell me about the last time your decision making skills as an evaluator were challenged?

During Sweep Two, I explained that the goal of the interview questions was to identify a critical incident that had occurred while he was evaluating an HIV/AIDS health education program implemented in Sub-Saharan Africa. The research participant asked me for the definition of a critical incident. I provided the following definition: In this study, a critical incident is an event or situation that influenced the outcome of an evaluation you were conducting. After I explained what type of evaluation experience I was looking for, I asked the research participant: "Can you think of a time when your skills as an HIV/AIDS program evaluator were really challenged? The participant then shared the following evaluation experience:

> We were evaluating all of the different health department programs [in a Sub-Saharan country] … [that] implemented many programs in HIV/AIDS health education [that were funded by international aid agencies]. We wanted to gain [insight from] a managerial point of view that would be used to provide input for a strategic plan … We [the research participant and another member of the evaluation team] conducted an interview with the HIV senior government management person, and she was quite hostile when I first met her. [Note: This is not a complete narrative; a summary of Case One's Narrative can be found in Chapter 4.]

Knowledge Representation

Sweep Three: Interviews One and Two. In this sweep, I used knowledge representation to generate rough drafts of the decision making timelines and narratives. I used the decision making timeline to help establish a chronological framework — a decision making timeline — for the narratives. During Sweep Three, I asked questions so that I could obtain clarification and eventually, validation of the decision making timeline. The decision making

timeline was used to create a "re-storyied" narrative (Creswell, 2007, p. 56) A re-storyied narrative was generated because field texts from interviewees do not generally adhere to a story sequence with a beginning, middle, and end (Creswell, 2007, p. 56). In addition, I did not include information in the narratives and decision making timeline that was not directly related to the event.

During interview one, I explained that the goal was to co-construct and verify the generated decision making timeline. Creation and refinement of the decision making timeline served as a tool for further understanding of the research participant evaluation experience [knowledge elicitation]. Sweep Three's purpose and interview questions are listed in Table 5.3.

Table 5.3

Sweep Three: Purpose and Interview Questions

Purpose	First and Second Interview Questions
Timeline Co-Construction, Verification, and Decision Identification	Where in the timeline should I put this?
	Do I have this right?
	Where in the timeline do you think this critical incident occurred?
	When do you feel that this decision had to be made?
	Where do you think that this decision resolved the critical incident?

In Sweep Three, I worded the questions to match the specific decisions that were being explored. For example, instead of asking the research participant, "Where in the timeline should I put this?" the research participant was asked, "Is it okay to begin the timeline by indicating that 'the evaluation was commissioned to evaluate the health department for Country X [country's

name]?" The research participant stated that it would be a good way to begin the timeline. He also clarified that (1) the government was required by the donors to conduct an external evaluation, and (2) he was asked to interview the HIV senior government department manager, who supervised the HIV Health Education programs. I continued the interview and asked the research participant, "When you approached the government official, the official wanted to know who you are and who sent you — is that correct?" He said, "Yes." He stated that her tone and actions indicated that there might have been a problem.

During Sweep Three, the telephone connection was lost due to a major storm at my location. An hour later, I was able to reestablish a phone connection with the research participant. It was decided that I should prepare a rough draft of the timeline and narrative. In addition, and that we decided that the narrative and timeline would be discussed during the second interview. The research participant also stated that he would be leaving for a month to conduct an evaluation in a rural area where telecommunication would be limited. He also stated that after his return to his home country, he would need time to write his evaluation report. Thus, we agreed to schedule the second interview a month and a half after the first interview.

I created a timeline and narrative based on the interview transcripts. Two weeks before the second interview, I sent electronic versions of the decision making timeline and the narrative rough draft to the research participant. The research participant confirmed that he had received the rough drafts and that he would read them before the second interview.

Sweep Four: Interview Two. Sweep Four was originally planned for the first interview. However, due to technical problems, the rest of the sweeps were conducted during the second interview. In addition, I explained that the goal of the second interview was to gain

further understanding of the decision making process. The purpose and interview questions of Sweep Four are listed in Table 5.4.

Table 5.4

Sweep Four: Purpose and Interview Questions

Purpose	Second Interview Questions
Deepening Understanding of Decision Making	What was the situation that let you know what was going to happen? What was it about the situation that let you know what to do? What led up to this decision? What were your concerns at this point? How would you summarize the situation at this point? What information did you use to make this decision? How did you get this information? What evaluation knowledge was necessary or helpful in this situation? What were your specific evaluation goals at this time? What were you hoping as an evaluator to accomplish at this point?

The second interview began with a review of the decision making timeline. I began the interview by providing the research participant an opportunity to review the rough draft of the decision making timeline. As the timeline was being reviewed, the research participant informed me that the initial interaction with the government official lasted about five minutes before he "pushed back." The research participant corrected my errors in the timeline. For example, when I stated, "Then you prepared to leave. You indicated that you didn't want to abuse authority and it was a delicate balance." The research participant replied, "I don't think I said that I wasn't going to abuse my power. It was what I thought to myself." The research participant corrected the decision making timeline when he said that he was not the "evaluation team leader." He stated that he was a "member of the evaluation team."

Corrections made to the timeline were applied to the narrative. When the research participant was reviewing the narrative, he stated that his decision to "push back" was portrayed as "too neutral." He also stated that the narrative needed to include that "his demeanor changed" and that he responded with "forceful tones" in his voice.

I asked the research participant, "What let you know what the interviewee was avoiding being interviewed?" The research participant replied that the official's response was unexpected. He stated:

> You know, when working with government officials, you tend to find three typical responses: Sometimes, there are very efficient and direct. They will answer with polite responses. They ask, "Can you be brief?" "What are your questions?" "Can we do it in thirty minutes?" No matter what they feel about the evaluation or the interview, they take it on. Other people are hesitant to participate. This may be due to many reasons. For example, they [the interview participants] may not feel up to speed or caught up. They may also wonder about the political context. They want to know if you are going to get them in trouble with someone. They may give you very short answers or very general answers. Then, there is a category of people who like to "charm you." They say, "I'd love to talk to you. What you are doing is very important! We are having a great success." I will often need to say, "That is okay. Back up. Tell me more. Did you actually *do* that [referring to claims of success.]?" They will often respond, "No, we are having great success" and they just keep on talking. That is typical of the "charming" types.

I asked the research participant, "What was it about the situation that let you know what to do?" The research participant stated that he relied on a previous evaluation experience to help him determine what to do. His response answered the questions, "What information did you use to make this decision? How did you get this information?" The research participant stated:

> In this situation, I had to be careful. I wasn't sure what I could or could not say. What helped me think this through was a previous evaluation experience. In that evaluation, government officials were also interviewed. It was the first time I had directly worked with senior government people. I was in a team that worked with a team leader with a great deal of experience. [Note: This is not the complete narrative; a summary of Case One's Narrative can be found in Chapter 4.]

The research participant spontaneously answered the interview question, “What evaluation knowledge was helpful in this situation?” when he stated:

> In evaluation, you need to learn how to negotiate the power differences. The state or the community being served should have the power to determine how money is spent and how the program is designed. Since the donors are funding the programs, they want a voice, too, and they want to see that the money is well spent. If not, they can take [the money] somewhere else.

The research participant shared some of his philosophical assumptions regarding the evaluation. For example, he said:

> You see, when you conduct an HIV evaluation, it is important to not only understand the nature of the evaluation, but you need to understand the evaluation context. In HIV evaluations, the time constraints are more pronounced. This is due to the overlapping and sometime conflicting processes between the donors and the HIV programmers. Many times the programmers don’t realize that they need to complete an evaluation before they will receive funding for the next program cycle. They don't realize that evaluations take a long time. The donor must remind programmers by saying, “The project cycle ends in six months and you need to complete the evaluation.”
>
> Compounding this problem is that HIV programs can also be highly funded and highly structured. What I mean by structures is that there are lot of institutions and lots of lines of reporting. There are a lot of coordinating bodies, advisory bodies, funding bodies, and lot of different players. The institutional environment can be complex with a lot of money floating through these structures.

When the research participant narrated his decisions and concerns, he answered the interview question, “What led up to that decision?” In his narrative, he stated:

> I felt that since the department manager was a public employee and I had been hired by the state to run an evaluation — on behalf of the donor — the manager needed to talk to me. … The manager started packing her briefcase … I also knew that we really needed the information. … I decided to let her know that not participating wasn’t a good idea. … I acknowledged that the manager was someone who wasn't appreciated or heard. … I reframed things. The manager’s hostility increased.
>
> I generally come across [as a] quite easy, non-confrontational person, but when I do get irritated, there is a clear change in character. ... It irritated me to hear the manager say, “I am important. I need to meet with the minister.” ... I believed this

> because the manager was a public servant and because the manager was managing the HIV department, and the department needed to address the scale of the HIV problem. They were quietly behind the whole "denialist" camp. The politics of [the view that HIV doesn't exist or that HIV does not cause AIDS] allowed them to make mistakes. I also know that they wasted money.

The research participant also answered the interview question, "What were your concerns at this point?" He stated that he asked himself the following questions:

> "Do you go over her head and make her even angrier? How do you work with someone who has quite reasonable concerns?" She probably was a very important and busy person. She probably was being harassed by a lot of people, but I couldn't change that, so we had to find some kind of way around it.

When the research participant introduced the first critical incident, he answered the interview questions, "What were your specific evaluation goals at this time? What were you hoping as an evaluator to accomplish at this point?" He stated:

> We scheduled an interview with the HIV senior government department manager to evaluate the HIV health department. The programs included HIV health education. A fellow evaluator and I walked into the manager's office for the interview…. I needed the information [from the interview for the evaluation report].

Sweep Five: Interview Two. In Sweep Five, I asked the research participant "What if" questions in order to obtain an understanding of the elements or factors that influenced the evaluator's decision making. The purpose and interview questions of Sweep Five are listed in Table 5.5.

Table 5.5

Sweep Five: Purpose and Interview Questions

Purpose	Interview Two Questions
Deepening Understanding of Decision Making	What alternatives did you consider?
	What would another evaluator possibly do differently?
	What other actions could you have taken?

	How would you have approached this decision earlier in your evaluation career? How would this incident have turned out differently if someone without your level of expertise had not been there? If you were to encounter this type of incident again, how would you approach it?

When I asked the interview question, "What alternatives did you consider?" the research participant stated, "If she continued to refuse to talk to me, I could have gone to her boss. I could have said, 'Well, the HIV Health Department Manager wouldn't talk to me. You sort it out.'"

I asked the research participant, "What would another evaluator possibly do differently?" the research participant stated, "I don't know what another evaluator would have done."

When I asked, "How would you have approached this decision earlier in your evaluation career?" the research participant replied, "Without that evaluation experience, I would probably been more reserved. I think I would have followed her basic strategy, which was to direct me to previous evaluation reports. That might have been easier."

Since the research participant stated that this experience occurred early in his evaluation career, I did not ask him, "How would this incident have turned out differently if someone without your level of expertise had not been there?" I forgot to ask the research participant, "If you were to encounter this type of incident again, how would you approach it?"

Sweep Six: Interview Two. In Sweep Six, a review of the narrative and timeline was further continued. I asked the research participant questions that helped clarify the narrative and timeline. The purpose and interview questions of Sweep Six are listed in Table 5.6.

Table 5.6

Sweep Six: Purpose and Interview Questions

Purpose	Interview Two Questions
Clarify and Validate Narrative/Timeline	(While summarizing the critical incident story, timeline, and reviewing the key factors) What other observations would you like to add regarding your decision making process? What other observations would you like to add regarding the critical incident(s)? Did I misunderstand this? Did I forget something? Is there something [sic] else you would like to add?

In Sweep Six, I found that the corrections to the documents provided more information about the critical incidents encountered by the research participant as well as his decision making. For example, when I asked, "What other observations would you like to add regarding your decision making process?" the research participant said, "If it had been another department – like the department in charge of diabetes – I would have responded differently to the hostile interviewee." He said that his response would have been different if the official had been in charge of health services for diabetics, as diabetes is not politicized and he had not conducted research on diabetes. He also stated that due to the politicized character of HIV, he knew that "everything he said would be politicized" and his extensive work in HIV "led him to feel that he had the authority to say something." He stated, "You [meaning me, the researcher] might want to add as something that influenced my [the research participant's] interaction with the evaluator."

During the sweep, I asked, "What other observations would you like to add regarding the critical incident?" The research participant's response provided an explanation as to why the evaluation participants may resist from being evaluated. The research participant stated that the

evaluation resistance occurs because the research participants tend to be over-researched and over-evaluated. Thus, they do not understand the evaluation process. He stated that (1) the evaluation participants need to meet with many people who are interested in conducting research on HIV/AIDS projects; (2) project monitoring and evaluation requirements could be "onerous" and could "duplicate a lot of effort"; and (3) stakeholders do not understand the amount of time it takes to conduct an evaluation.

The research participant explained that because of the political climate, he "decided not to recommend a departmental reorganization [in the evaluation report] because it [department reorganization] was likely not to happen." He explained that his recommendations were based on his evaluation findings. His findings indicated that the "prevention programs were quite fragmented and uncoordinated." For example, he identified the following prevention programs as "not having any sense of urgency attached to them:" ARV [Antiretroviral Drug] program, HIV health education workshops, condom distribution program, and mass awareness campaign. The evaluation recommendations, in the evaluation report, stated that the HIV/AIDS health programs should be "driven by a central person" and that the programs should be coordinated with the other health departments.

At the end of the second interview, I asked the research participant, "Did I forget something?", and he stated, "No, I think you deconstructed things very well." I also asked, "Is there something [sic] you would like to add?" His response was, "No, you captured the story. I don't think there are any big points that you are missing." At the end of Sweep Six, I thanked the research participant for his participation in the research study and concluded the interview. The decision making timeline and narratives were finalized based on the knowledge gained from the second interview.

Decision Making Timeline. A decision making timeline was generated for each case, depicting the following: evaluation context, the critical incidents the evaluator faced, events that occurred, and the decisions that were made. The decision making timeline provided a high level overview of what occurred in the research participant's story.

The decision making timeline design consisted of three sections: Evaluator, Evaluation Context/Stakeholder, and Event Timeline. The Event Timeline depicts the general time span of the story. The Evaluator section indicates the evaluator's actions, motivations, goals, theory(ies), reflections, and decisions. The Evaluation Context/Stakeholder section indicates the events based on the following: (1) Evaluation context or stakeholder's actions, (2) Motivations, (3) Goals, (4) Theory(ies), (5) Reflections, and (6) Decisions.

In the timeline, boxes and arrows are used to guide the reader. Critical incidents are labeled "CI" and the decisions are labeled "Decided" in bold font. Critical incidents are indicated by an arrow with a dashed line and are encased in a double-lined box. Decisions are indicated by a solid black arrow and are encased in a box labeled "Decided." An unfilled arrow points to boxes that provide information about the story's context, as well as the evaluator's and stakeholders' goals.

I used the rough drafts of the decision making timelines as an interview tool, for knowledge elicitation, and to develop a final version of the decision making timeline. An example of a rough draft of the decision making timeline is provided in Figure 5.1. Embedded in Figure 5.1 are questions I asked to obtain further clarification. For example, the research participant was asked, "How many months into the evaluation was the interview scheduled?" He answered that the interview was scheduled two weeks into the implementation of the evaluation.

He was also asked, “How long was the conversation that focused on encouraging the government official’s participation?” The research participant reported that it lasted five minutes.

It should be noted that Case One was my first case. As a result, Case One’s decision making timeline became a guide for the other six cases’ decision making timelines. For example, when the research participant requested assistance in understanding the decision making timeline, I decided to add “CI” as a label to identify the critical incidents in the timeline.

The second interview resulted in a finalized decision making timeline. Similar to the rough draft, the finalized decision making timeline was divided into three sections. An example of the final decision making timeline based on Case One is provided in Figure 5.2.

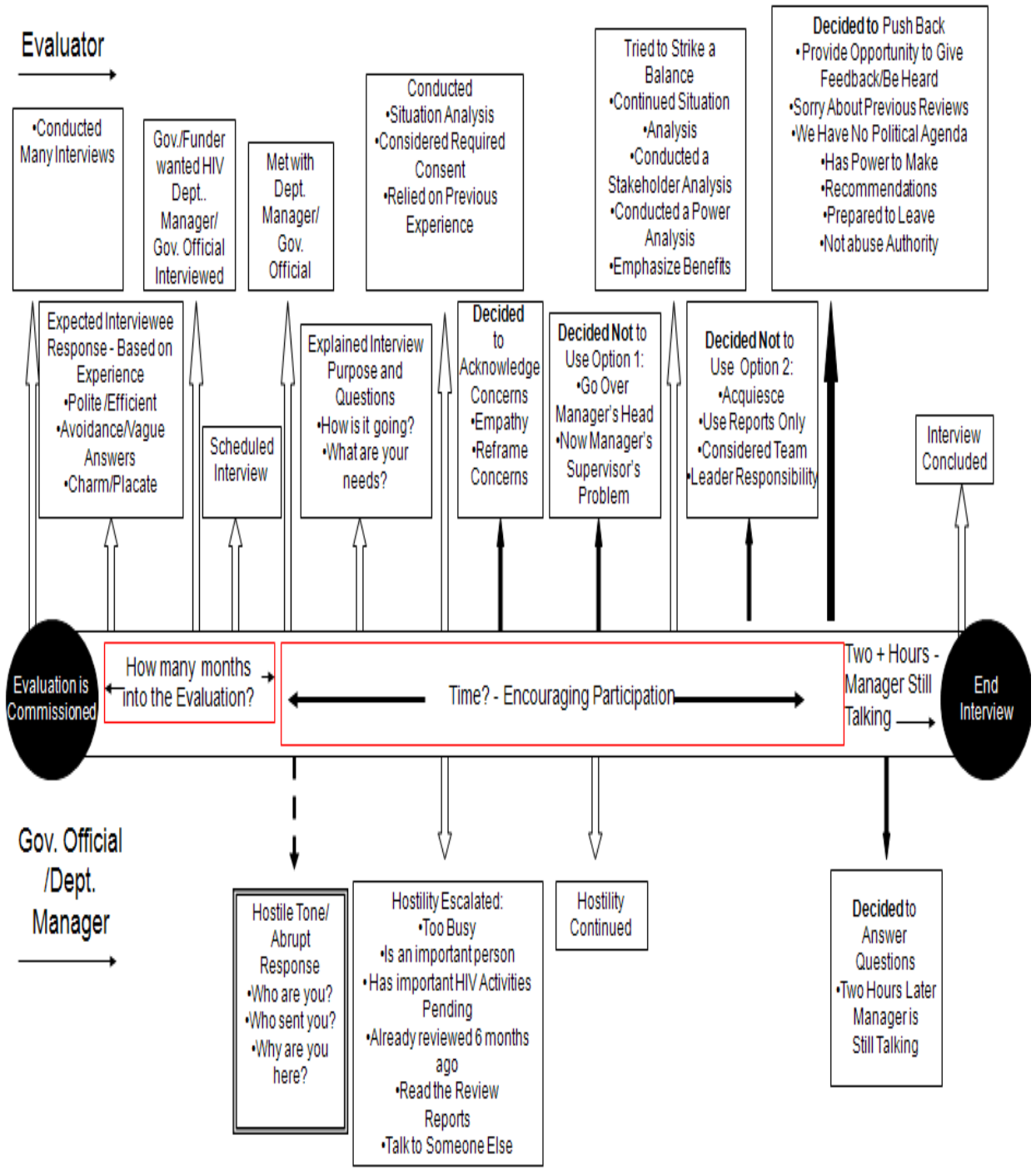

Figure 5.1. Case One: Rough draft decision making timeline with interview questions.

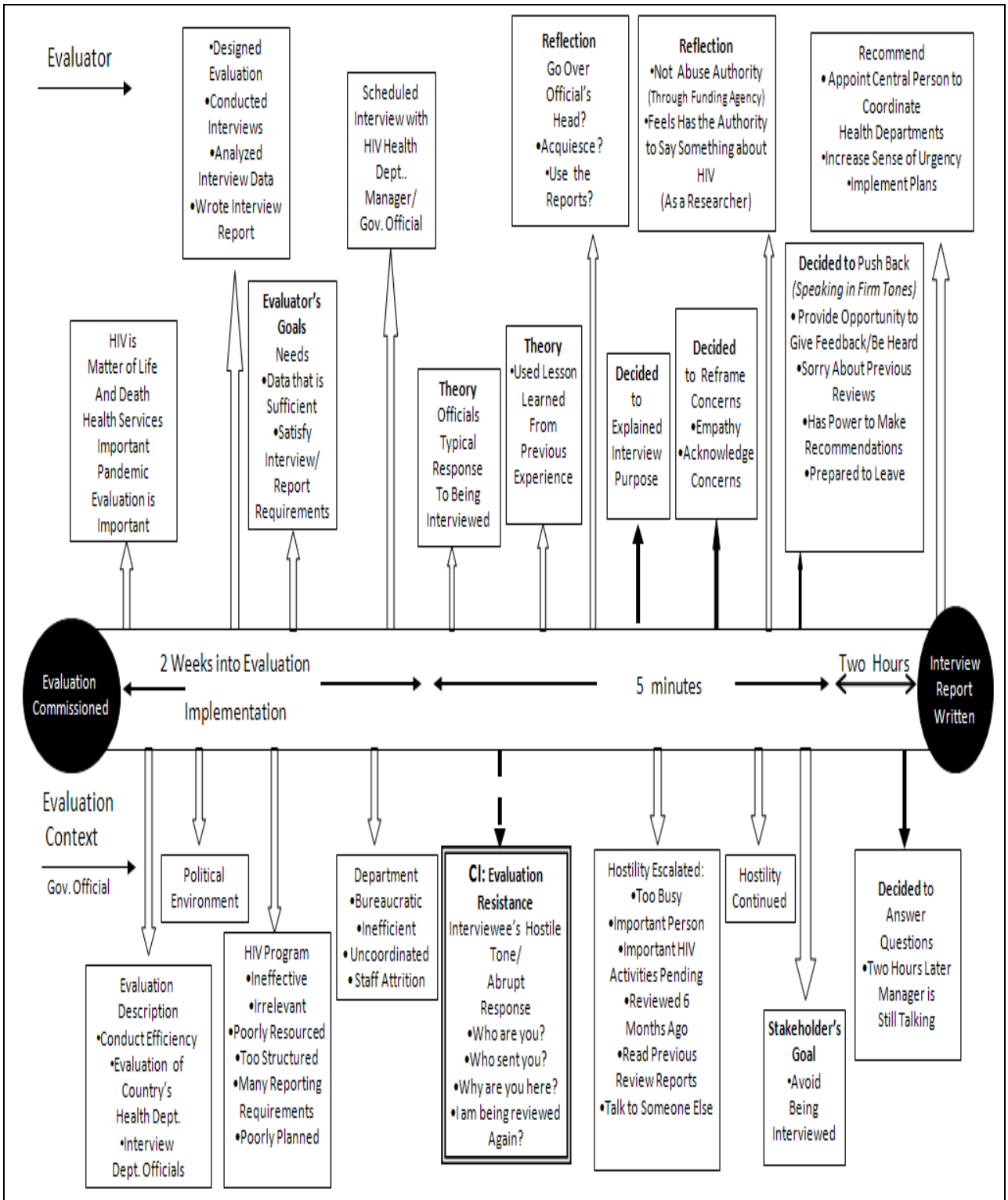

Figure 5.2. Case One: Finalized evaluation resistance decision making timeline.

The decision making timeline depicted the time span of the evaluation story and provided the following information:

(1) The story began when the evaluation was commissioned.

(2) The interview with the official in the health department was scheduled within a two-week timeframe.

(3) The initial interaction with the official spanned five minutes.

(4) The official decided to answer the questions and continued talking for at least two hours after the critical incident was resolved.

(5) The evaluation story ended when the evaluation report was written.

The Evaluator section provided the following information:

(1) The motivations for conducting the evaluation were: the evaluator considers HIV a matter of life and death, health services are important, HIV has reached pandemic proportions, and evaluation is important.

(2) The evaluator designed the evaluation, conducted interviews, analyzed interview data, and wrote the interview report.

(3) The evaluator's goals were to gather data that were sufficient and satisfactorily met the interview report requirements.

(4) The evaluator scheduled an interview with the HIV Health Department.

(5) The evaluator expected a typical department official's response to an interview request.

(6) After experiencing the critical incident, the evaluator utilized a theory on how to approach the hostile interviewee, based on a previous experience.

(7) While the evaluator was contemplating a solution to the critical incident, he reflected on whether he should:

(a) take his concerns to the official's supervisor

(b) take the interviewee's suggestions,

(c) in particular, just read the reports instead of conducting the interview.

(8) The evaluator decided to explain the purpose of the interview.

(9) When that did not work, he decided to appeal to the interest of the interviewee and show empathy as a way of convincing the government official to participate in the interview process.

(10) When this did not resolve the critical incident, the evaluator reflected on his authority as:

(a) a consultant hired by the funding agency

(b) a researcher to speak on HIV.

(11) The evaluator decided to push back — speaking in firm tones — and indicated that participating in the interview would provide the official an opportunity to provide feedback, indicate that he was sorry that she had just been interviewed, and indicate that as an external evaluator he had the power to present her concerns to the funding agency.

(12) After the department official decided to participate in the interview and all of the interviews for the evaluation were completed, the evaluator recommended that:

(a) The department should appoint a central person to coordinate all of the HIV programs

(b) There should be an increased urgency in the delivery of HIV health services,

(c) The department should implement the existing plans.

The Evaluation Context/Stakeholder section provided the following information:

(1) Included a description of the evaluation

(2) Stated that the evaluation was conducted in a political environment

(3) Declared that the HIV program was ineffective, irrelevant, poorly resourced, had too many reporting requirements, and was poorly planned

(4) Observed that the health services department was too bureaucratic, inefficient, uncoordinated, and experiencing staff attrition

(5) Identified the critical incident as evaluation resistance

(6) Indicated the interviewee's continued hostility

(7) Identified the stakeholder's goals

(8) Identified the stakeholder's decision to participate in the interview process.

Narratives. Based on the transcripts and timeline, I generated "re-storyied" narratives. The research participants reviewed the narratives to ensure that they were accurate. During co-construction, the research participants' stories were split into two narratives. The first narrative focused on the critical incidents that the expert evaluators shared. The second narrative focused on the evaluators' philosophical assumptions. The two narratives helped me to distinguish which part of the research participants' stories focused on the evaluators' philosophical assumptions and which parts focused on the critical incidents.

Data Analysis

Sweep 7: Interview Two. Sweep Seven focused on data analysis. Sweep seven followed Creswell's (2007) recommendation that the narrative analysis should be utilized to capture "detailed stories or life experiences of a single life or the lives of a small number of individuals" (p.55). In a narrative analysis, the codes, categories, and themes are based on the

stories that the research participants share (Creswell, 2007). During the analysis, the level of abstraction increases as the codes move from categories to themes (Saldana, 2009).

During and between the interviews, I used the research participants' stories to generate a narrative as well as a decision making timeline. I used the interview field notes and the rough drafts to assist me in creating the narratives and timelines. A rough draft of the decision making timeline and narrative was developed after the first interview. When I created the narratives and timelines, my insight or understanding of the research participant's experience was unavoidably interwoven into the documents.

I generated memos so that I could to reflect on my "behavior and thoughts as well as the phenomenon under study" (Glesne & Peskin, 1992, p. iii). The memos also served as part of the study's audit trail. Figures 5.3 and 5.4 provide examples of the memos I used to reflect on my behaviors, emotions, and values. The figures also provide an opportunity to reflect on the research participants' behaviors, emotions, and values. Figure 5.3 reveals the research participant's frustration, despair, and anger due to inadequate HIV health services. Further, it also explains why I decided not to interview the recipients of HIV health education services. This is consistent with Patton's statement that qualitative research is "time consuming, intimate, and intense" (2002, p. 35). Figure 5.4 provides an insight into the influence that politics had on the research participant and the stakeholders.

I could hear the research participant's frustration, despair, and anger in his voice. I tried to not get upset during this part of the story. I had specifically tried to conduct one-off HIV/AIDS research because I knew that the topic was sensitive. For example, I did not interview HIV/AIDS patients about what they thought about the health education services programs provided — it would have been too upsetting.

Kvale (2007) stated that "the knowledge produced in a research interview is constituted by the interaction itself" and that the "interaction may also be anxiety provoking" (p.4). I could hear the evaluator's emotions swell in this part of the story. During the review of the narrative, the research participant said that I caught his voice and that reading the narrative had been an enriching experience. It was an enriching experience for me as well.

During the first interview, I reflected on how I was only hearing the evaluator's side of the story. If my research design had included interviewing an evaluator and an evaluation participant in the same health education evaluation, I would have found richer data. Upon reflection, I observed that it would have been quite difficult to arrange interviews with the second participant.

Figure 5.3. Case One memo: Politics, evaluator's frustration, and researcher's reflections.

"People are suspicious of your political agendas. Adding to this are HIV related political agendas. In this evaluation, the country government officials wouldn't actually state their position on HIV and AIDS. They denied that they were 'denialists' — people who deny that HIV caused AIDS — but their actions indicated differently. That was translated to people in the department taking up that [denialist] camp and not wanting to provide ARV's [Antiretroviral drugs], or they were providing alternatives that weren't really working … So [when conducting an evaluation] you were always trying to figure out what somebody's real position was ... You [as an evaluator] become a player in all of that drama. It was a running challenge throughout the evaluation."

This quote from Case One describes the effect politics has on evaluation stakeholders. The government's politics affected the departments, which in turn affected the types of

interventions and services the department was implementing. The research participant noted later in the narrative that politics influenced whether department officials would participate in HIV/AIDS evaluation activities—including interviews. The research participant indicated that he was influenced by the political environment as well.

Figure 5.4. Case One memo: Impact of politics on stakeholders, occurring in the phenomenon under study.

After the interviews were completed, I conducted a stand-alone data analysis of the decision making timelines and narratives. My first goal was to identify the critical incidents and the decisions in the stories. Figure 5.5 illustrates how I conducted my initial data analysis for Case One and used the MS Word comments feature to highlight the possible codes, categories, or list some of the questions that I had planned to ask during the second interview.

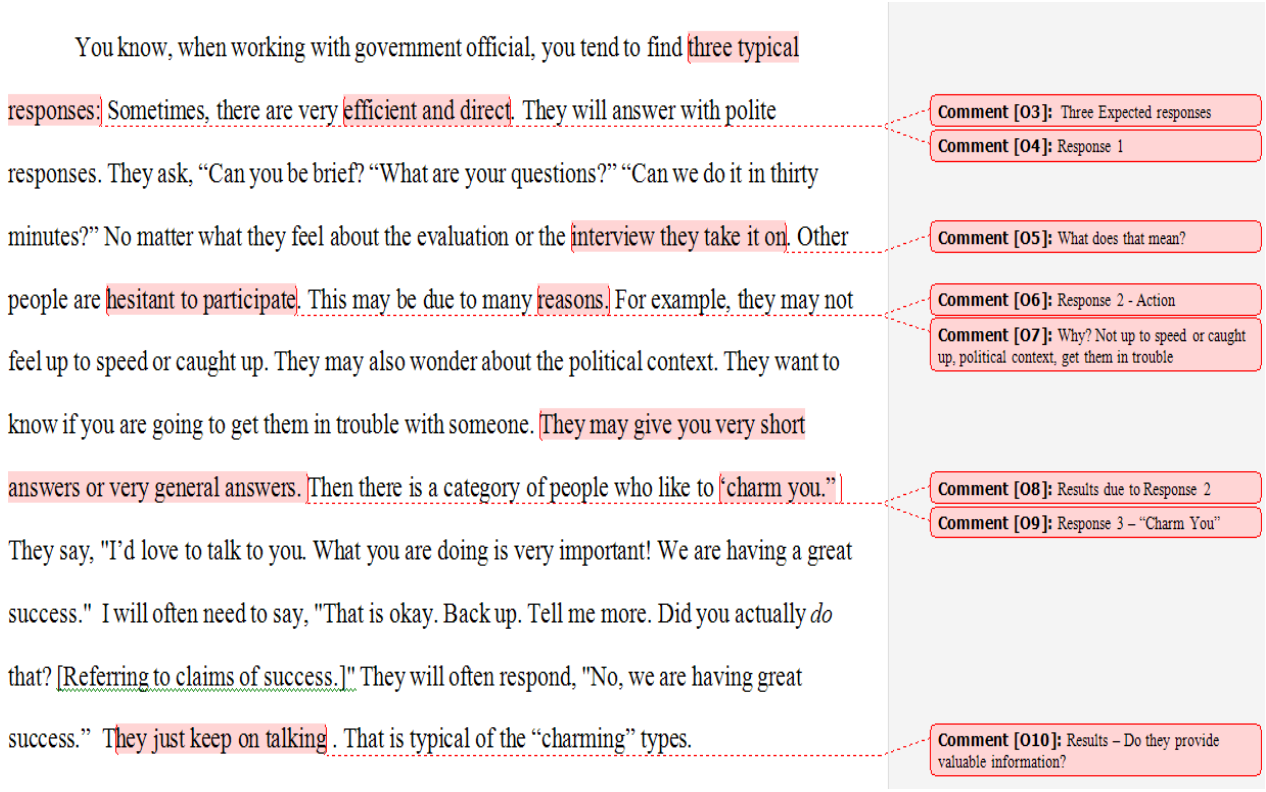

Figure 5.5. Initial stand-alone data analysis

As data analysis continued, tables for each case were created that listed the following: (1) the critical incidents, (2) decisions made in response to the critical incidents, and (3) quotes that depicted the critical incidents and decisions. An example of the tables used for each case is provided, based on Case One (See Table 5.7).

Table 5.7

Case One: Critical Incident, Evaluator's Decisions, and Quotes

Critical Incident (CI) and Evaluator's Decisions	Quotes
CI: Evaluation Resistance	"She was quite hostile when I first met her."
Decision 1: Evaluator Explains the Purpose of the Interview	[The evaluator] "decided to share the purpose of the interview."
Decision 2: Evaluator Reframes Stakeholder's Concerns and Expresses Empathy	The evaluator decided to say to the official, "This [interview and evaluation] is an opportunity for an outside person to say what you want to them to say … If you think you have something important to say that hasn't been heard, then go for it."
Decision 3: Evaluator Confronts Stakeholder	The evaluator indicated that, as he prepared to leave, he said, "It is up to you."

In Table 5.7, the first column is divided into two sections: Critical Incidents — indicated by the initials "CI" — and Decisions, which are numbered. The second column lists the quotes from the narratives. The quotes do not provide in-depth details about the critical incident or decisions; instead, they indicate the beginning of the critical incident and the point at which decisions were made.

During data analysis, my other goal was to identify the factors that influenced the expert evaluator's decision making for each case. Thus, during data analysis, I created a table that alphabetically listed the factors that I identified in the narratives with their corresponding quote(s). Based on Case One, Table 5.8 provides an example of the tables I generated for each

of the cases. In the first column, the factors influencing the evaluator's decision making are listed. At times, the factors were further divided. For example, in the first row, the Accountability factor is divided into three sections: Evaluator, Organizational, and Program Manager. However, the Bureaucracy factor is not divided as it only refers to the HIV health department's bureaucracy. The second column lists the quotes that illustrate how the factors appeared in the narratives.

Table 5.8

Case One: Factors that Influenced the Evaluator's Decision Making

Factor		Quote
Accountability		
	Evaluator	"Evaluators are accountable to the evaluation sponsor and, in this case, to the evaluation team leader ..."
	Organizational	"When the government or a program takes money from anyone, it becomes their responsibility to spend it wisely."
	Program Manager	"The department manager was a public employee, and I had been hired by the state to run an evaluation; the manager needed to talk to me."
Attitudinal		
	Evaluator	"I generally come across [as a] quite easy, non-confrontational person, but when I do get irritated, there is a clear change in character."
	Program Manager	"The manager was quite hostile."
Bureaucracy		"Big organizations like HIV health departments have many hierarchal layers. … I had seen that the department was a mess. I felt I had a right to confront the manager and say, 'This isn't good enough.'"
Civil Unrest		"The situation in the department [was that] they need consultants to do a lot of their work. It was the result of the transition from decades of civil upheaval."
Complexity		"Compounding this problem is that HIV programs can also be highly funded and highly structured. … The institutional environment can be complex with a lot of money floating through these structures."
Conflict		
	Evaluator	"I felt I had a right to confront the manager and say, 'This isn't good enough.'"

Factor		Quote
	Program Manager	"It might be why the manager was resisting the interview."
Consent		"Obtaining consent for a research project can be different from obtaining consent in an evaluation."
Context		"You see, when you conduct an HIV evaluation, it is important to not only to understand the nature of the evaluation, but you need to understand the evaluation context…"
Control		
	Emotional	[Evaluator: transitions from easy going to confrontational] [Program Manager: transitions from hostile to accommodating]
	Locus of Control	[Evaluator: Accepts that there are some elements in the situation that can't be changed but believes that he can influence the evaluator to participate] "… we need to find a way to work around it."
	Cooperation	"Evaluation is not always collaborative but [it] needs stakeholder cooperation."
Data Quality		"I think I would have followed her basic strategy, which was to direct me to previous evaluation reports. That might have been easier. Still, I wouldn't be happy to have weak information."
Experience		
	Interview	"When working with government official, you tend to find three typical responses ..."
	Evaluation	"What helped me think this through was a previous evaluation experience. … Without that evaluation experience; I would probably been more reserved."
	Field	"I ... have enough work experience. … I can speak with some authority."
Evaluator's Role		"You know, we [external evaluators] have a lot of power as outside people."
Goals		
	Evaluator	[Conducted Interview for] "data gathering"
Organizational		"In this evaluation, the government was fulfilling donor requirements by commissioning an evaluation of the health departments, which included HIV health education programs. The government wanted an independent evaluation."
High Stakes		"I think because my reaction to it — the devastation due to HIV … my interaction with the department manager was influenced by what I felt was at stake."

Factor	Quote
Multiple Stakeholders	"You have to work with several people, so it is important to triangulate among yourselves."
Over-evaluated	"[Evaluation] participants tend to be over-evaluated … [For example, the manager told me that there had been a review process six months ago."
Planning	"There is all this time put into planning, but not practices. There are various reasons it doesn't translate into practice."
Politics	
Organizational	"The department was taking up the [political 'denialist'] camp and not wanting [sic] to provide ARVs or they were providing alternatives that weren't really working.
Evaluator's	"I think it was the political environment and the scale of the HIV/AIDS problem [that] influenced my reactions."
Power	
Evaluator	"As an evaluator, you don't want to overplay the authority given to you by the evaluation funders … You know, we [external evaluators] have a lot of power as outside people."
Manager	"The manager said, 'I can't be interrupted by a low-level consultant.'"
Program Stakeholders	"The state or the community being served should have the power to determine how money is spent and how the program is designed."
Program Donors or Funding Agencies	"The donors are funding the programs; they want a voice too, and they want to see that the money is well spent. If not, they can take it [the money] somewhere else."
Time	
Manager	[Evaluator's reflection] "Maybe she [the manager] was too busy."
Evaluation Timelines	"In HIV evaluations, the time constraints are more pronounced. … Donors require a crazy compressed timeline, but they also need to have the evaluation completed."
Evaluator	"There is another difficulty in this quick evaluation work … You need to take into consideration that time is so limited."
Timing	"Often you schedule an interview two to three weeks in advance. Then people don't show up, so you need to reschedule the interview and [yet you still need to] meet the targeted [evaluation] timeline.
Training	"I overcame the challenges [the interviewee' attempt to avoid being interviewed] by relying on previous experience and training."

Factor	Quote
Utilization	"I [the evaluator] considered my responsibility more than parrot[ing] what [the official] said. My job was to make judgments on what people say and make a recommendation that has an effect on the delivery of HIV health services."

Case One is utilized as an example of an expert evaluator's situation analysis and decision making process. The evaluator exhibited the four recognition by-products (Klein, 1997) that are present when the person conducting a situation assessment has identified a prototype of a typical action or response in a situation. The four recognition by-products are identified as expectancies, plausible goals, typical actions, and relevant clues. The research participant reported his thought process through his reflections on the situation at hand.

Kundin's (2009, 2009) research provided further clarification on how evaluators make practice decisions when conducting evaluations. Kundin indicated that evaluators utilized theories-in-use in real world practice, which are informed by practical reasoning, and reflection-in-action. Kundin provides a classification of the elements that influence practical reasoning, namely, working logic, general logic, and logics of action.

Table 5.9 depicts that during a situation assessment the evaluators (1) utilized theories-in-use in the real-world, (2) judgments, (3) expectations, (4) actions, (5) logic, (6) reflections, and (7) decisions. Table 5.9 also lists the (1) goals, (2) results, and (3) relevant cues used in the decision making process. The elements are depicted in the table, as far as possible, in the same order as they appeared in the narrative.

In Table 5.9, the first column lists the elements used during a situation assessment or a decision. The second column reports the quotes that match with the elements. There are four

table spanners embedded in the table: Evaluation Context, Decision 1, Decision 2, and Decision 3. The rows contain specific information about the elements listed under the table spanners.

Table 5.9

Case One Situation Assessment Based on Evaluation Resistance

Element	Quotes
	Evaluation Context
General Logic: Schedule Interview and Evaluate HIV Health Department	"We [the evaluator and another member of the evaluation team] scheduled an interview with the senior government department manager to evaluate the HIV health department."
Plausible Goals: Gather Data	"My [the evaluator's] goal was to gather data that met the interview report requirements."
Theories-in-Use in Real World: Three Typical Interview Responses	"[there are] three typical [government official interview] responses: Sometimes, they are very efficient and direct...Other people are hesitant to participate … Then, there is a category of people who like to 'charm you.'"
Relevant Cue: Abrupt Manner	"The manager said in an abrupt manner, 'Who are you? What do you want? Who sent you?"
Theories-in-Use in Real World: Judgment that the Interviewee is Hostile	"The manager was quite hostile."
	Decision 1: Evaluator Explains the Purpose of the Interview
Relevant Cue: Hostility Continued	"The hostility continued."
Theories-in-Use in Real World: Relied on Previous Evaluation Experience	"The lead evaluator was always respectful. He acknowledged that the premier was an important political figure … that we needed to provide this information to funders …"
Reflection-in-Action: Acknowledge Manager's Concerns and Empathize	"Maybe she [the manager] was too busy."
Working-logic: Report Hostile Interviewee to Her Manager	"I could have gone to her boss."
Reflection-in-Action: Further Anger the Manager	I had to ask myself, "Do you go over the manager's head and end up making the manager angrier?"
Reflection-in-Action:	"As an evaluator, [I thought to myself] you don't want to over play the authority given to you by the funders … It can

Element	Quotes
Negotiating Power Differences	become abusive."
Decision 2: Evaluator is Respectful, Reframes Stakeholder's Concerns, and Expresses Empathy	
Theories-in-Use in Real World: Judgment that Hostility Continues	"The manager's hostility increased."
Decision 3: Evaluator Confronts Stakeholder with Ultimatum	
Working-Logic: Evaluator has Power	"You know, we [external evaluators] have a lot of power as outside people. We are not caught up in the politics of all of this. We are able to make recommendations."
Result	She decided to participate in the interview, and two hours later, she was still talking.

My next focus in the study was to develop a model for each of the cases. I conducted a stand-alone analysis of the case narrative and decision making timeline when creating the model. The models were divided into five sections: Evaluation Context, Critical Incident(s), Situation Analysis, Satisficing, and the Decision Making Process.

A model based on Case One is provided in Figure 5.6. The first box, titled Context, provides a description of the evaluation. The description lists the type of evaluation [in this case a formative evaluation] as well as the evaluand. The second section, titled Critical Incident indicates what type of critical incident occurred: Evaluation Resistance [during an attempt to interview a department official]. The next section, "Situation Assessment," indicates that during an interview with a government official the evaluator conducted a comparison between what was occurring and what was expected.

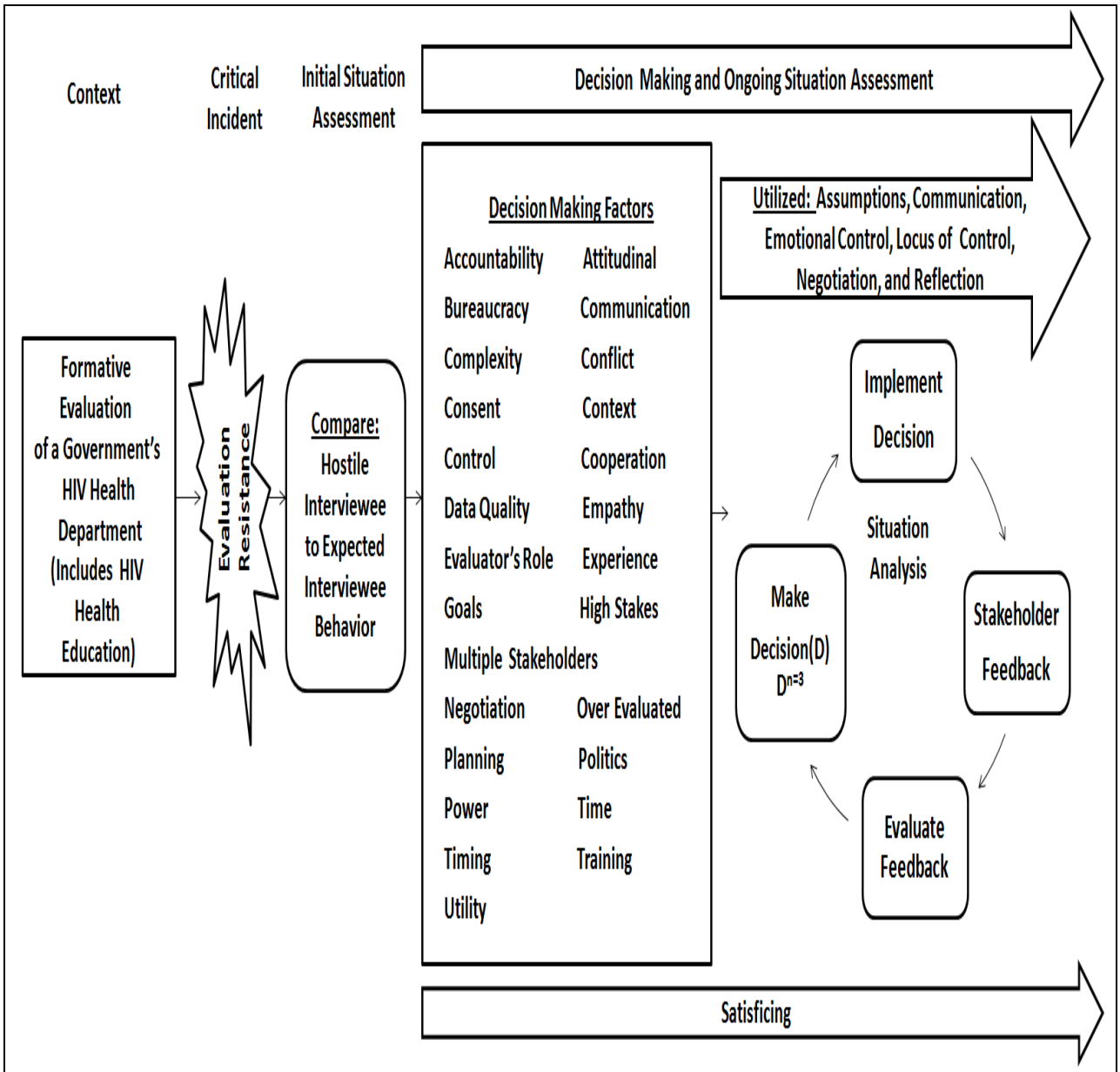

Figure 5.6. Case One Model: Evaluation Context, Critical Incident, Situation Assessment, and Decision Making Process.

The fourth section in the model lists the factors that influenced the evaluator's decision making, indicates that the evaluator utilized satisficing, and describes the evaluator's decision making process. My initial data analysis indicated that 26 factors influenced the evaluator's decision making. The 26 factors are listed in the box labeled, "Decision Making Factors." In addition, data analysis revealed that satisficing occurred because—as the evaluator indicated—(1) the interviewee was busy, overworked, and over evaluated, (2) the interviewee's initial situation was not a preferred interviewee response, and (3) in order to gather strong data the evaluator needed to settle for a less than optimal interview — a contentious interview in which the interviewee finally agreed to participate and where the interviewee talked for two hours. The solution was less than optimal because the evaluator implied that he would have preferred a professional interaction and a brief interview that was to the point.

The model also demonstrates that the research participant utilized the following: Assumptions, Reflection, Emotional Control, Locus of Control [later changed to self-determination], Negotiation, Empathy, and Communication to assist him in the decision making process. The narrative indicates that the evaluator utilized reflection-in-action before responding to the hostile interviewee. An example of reflection-in-action occurred when the research participant stated, "I asked myself, 'how do you work with someone who has quite reasonable concerns?'"

In one instance, the evaluator indicated that he utilized emotional control to assist him throughout the situations encountered. The evaluator stated, "I generally come across [as a] quite easy, non-confrontational person." This statement indicated that the evaluator had been trying to keep his emotions under control. However, eventually the evaluator became irritated. The research participant stated, "When I do get irritated, there is a clear change in character."

The evaluator used interpersonal skills to convince the interviewee to cooperate. The interpersonal skills the evaluator used were Negotiation, Empathy, and Communication. For example, the research participant stated, “In evaluation you need to learn how to negotiate the power differences.” The use of empathy was indicated when he stated, “This way I acknowledged that the manager was someone who wasn't appreciated or heard.” The evaluator used communication skills to defuse the situation. For example, he stated, “I tried to explain the purpose of the interview,” and “I reframed things.”

In the model, I labeled a section “The Decision Making Cycle.” The Decision Making Cycle indicates that decision making is responsive to the feedback provided in the environment. During the decision making cycle there are action/feedback loops that are informed by repeated situation assessments. I used arrows in the model to indicate that the decision making cycle was iterative and dynamic. Further, I used boxes to indicate the major steps in the decision making cycle: Generated Decisions, Make Decision, Implement Decision, Stakeholder Feedback, and Evaluate Feedback. In the box labeled Generated Decision, $D^{n=3}$ indicates that three decisions were made.

Phenomenography

I was not able to provide an example of how phenomenography was used in this research based on one case. To conduct phenomenographic research I needed to conduct a comparison of all of the cases. When I compared the cases, I examined them to determine if there were any patterns that would help me identify variations between the critical incidents, the factors the influenced the evaluators’ decision making and how the evaluators made decisions. I found (See Chapter 7) that the critical incidents could occur as barriers to or facilitated the implementation of an evaluation. Further, I also found that the critical incidents were either expected or

unexpected. I also found that the evaluators reported factors as either positive or negative factors that influenced their naturalistic decision making.

When I analyzed how the evaluators made decisions, I found that the evaluators generated a solution based on an initial situation assessment. Then evaluators tested the solution, and gathered stakeholder feedback; however, if the solution did not work, they continued this process [situation assessment, generate a solution, test the solution, receive feedback] until a solution was achieved. Alternatively, if the evaluator did not receive feedback, the evaluator conducted a situation assessment and then relied on his or her judgment to decide on how to proceed. In addition, if the evaluator felt that the choices were limited, the evaluator decided to work within the constraints they encountered.

Chapter Summary

This chapter provides an in-depth description of the protocol that I utilized to conduct the interviews and data analysis for my research. Case One was used to provide a description of how CDM and Narrative Analysis were used in my study. Examples of how the research questions were answered are also provided. Phenomenography guided my research. It should be mentioned that a single case could not be used to describe how phenomenography was used in this study. Nevertheless, a general description of how phenomenography was used in this research is provided. The next chapter, Chapter Six, provides the results derived from the seven cases.

CHAPTER 6

RESULTS

This chapter presents a description of the results obtained in the current study. The results are presented according to the three research questions. The first section of this chapter reports the critical incidents that the expert evaluators encountered. The second section reports the factors that influenced the expert evaluators' naturalistic decision making. Finally, the third section reports how the expert evaluators made naturalistic decisions. The results in each section of the chapter are listed in alphabetical order and organized by the question's theme, categories, and sub-categories — later referred to as types. Quotes from the narratives supporting the results are also provided.

Research Question One:

What critical incidents do expert evaluators encounter during evaluations of HIV/AIDS health education programs?

This section provides the results of the data analysis, which led to the categorization and identification of the critical incidents the expert evaluators encountered. Tables and narratives are used to report the results. This section follows a general pattern: (1) definition of the categories; (2) quotes that support the findings or a table for reporting the results when there was more than one type of critical incident; (3) brief narrative summaries of the types of critical incidents; and (4) the literature related to the results.

Data analysis yielded the themes, categories, and sub-categories, which were utilized to depict the critical incidents the expert evaluators encountered. Based on the data analysis, the

categories were identified as follows: Barriers to Communicating with Stakeholders, Barriers and Enablers to Community Entry, Barriers to Coordination, Barriers to Understanding the Evaluand's Context, Budget Constraints, Evaluation Resistance, Successful Collaboration with Stakeholders, Threats to High Quality Data, Threats to the Evaluation Design, and Time Constraints. The sub-categories were listed as the types of critical incidents that occurred in each category.

Barriers to Communicating with Stakeholders

Communication is "Any act by which one person gives to or receives from another person information about that person's needs, desires, perceptions, knowledge, or affective states" (National Joint Committee for the Communicative Needs of Persons with Severe Disabilities, 1992, p. 3). The barriers to communication were the following: Geographical Barriers, Language and Culture, and Non-Communicative Co-PIs and a Domineering PI (See Table 6.1).

Table 6.1

Barriers to Communicating with Stakeholders

	Critical Incident
Barriers to Communicating with Stakeholders	Case Number and Quote
Geographical Barriers	Case 7: The evaluator stated that "throughout the evaluation process we communicated through email and met via Internet. …. It was a way to bridge over geographical gaps."
Language and Culture	Case 7: The evaluator explained that, "The nurses explained to us that some words do not exist in their local language ….. and that some words have a more powerful meaning in their local language."
Language and Culture	Case 4: "We [the evaluation team] wanted people that spoke English and at least one of the two local languages spoken in that country."

Critical Incident	
Barriers to Communicating with Stakeholders	Case Number and Quote
Non-Communicative Co-PIs and a Domineering PI	Case 6: The evaluator stated that the team "had established a system of having weekly calls … It was during one of those calls that we noticed that the PI [Primary Investigator] was the only person talking …. The PI was speaking on behalf of everybody."

According to Owen (2007, p. 121), "dissemination [of information] involves strategies and channels designed to inform audiences about the relevant aspects of an evaluation … For this to occur, channels of communication must remain open throughout." In Cases Four, Six, and Seven, barriers to communicating with stakeholders were encountered. The barriers were a result of geographical barriers, challenges due to different languages and cultures present in the evaluation context, and non-communicative stakeholders.

Communication challenges can occur when conducting programs in which stakeholders are geographically distant. Spooner, Flaxman, and Murray (2008) indicated that technologies such as telephone and email could be utilized to assist in overcoming geographical barriers when stakeholders are geographically distant. In Case Seven, telecommunication was used to compensate for the "geographical gap" the evaluation team encountered.

In Cases Four and Seven, the evaluation teams encountered language and cultural barriers. According to Ebbutt (1998), culture is a:

> constellation of both written and unwritten expectations, values, norms, rules, laws, artifacts, rituals and behaviours that permeate a society and influence how people behave socially." Language is intertwined with culture and can become a barrier to communicating with stakeholders if languages and culture differ. (p. 416)

In Case Four, the evaluation team too actions to overcame language and cultural barriers. They utilized interviewers who were from communities in which the evaluation was being

conducted or from similar communities. The evaluation team hired interviewers who spoke English and at least one of the two local languages, as a way to overcome the language barrier.

Participatory evaluation is one of the major approaches used for evaluating international development programs (Bamberger, 2000). In participatory evaluations, stakeholders are intrinsically involved in the evaluation process (Bamberger, 2000). In Case Six, the primary investigator dominated the conversation and created a hostile work environment. The evaluator thus used an investigative strategy to open the channels of communication. The evaluator's extensive evaluation experience in conducting evaluations in different organizations and cultures assisted in framing the investigative questions such that the primary investigator would not be offended.

Bulmer and Warkwik (1993) and Smith (1990) recommended that evaluators address cultural challenges when conducting evaluations in developing countries or environments in which the stakeholders' cultures or languages differ from those of the evaluators. Bulmer and Warkwik (1993) emphasized the importance of adapting data gathering instruments so that they are sensitive to cultural and linguistic differences. This is especially necessary if the instruments were originally developed in another cultural context. In Case Seven, a previous international evaluation experience informed the decision on how to address the differences in culture and language. It also contributed to ensuring that the data collection instruments, developed for a similar evaluation, were modified to the current evaluand's context. A local team was hired to assist in addressing the lexical and conceptual language differences between the evaluation team and the evaluation participants.

Barriers and Enablers to Community Entry. Community entry is achieved through the actions taken "to gain entry into the world of the people who experience the issue being

studied ... [where entry is a] systematically planned encounter between researchers and the community collaborators during each stage of research [evaluation]" (RTI, 2004, p. 25), and those instances in which partnerships are based on trust (Mertens, 2008). Three subcategories emerged under the category of Barriers and Enablers to Community Entry (See Table 6.2). The sub-categories are Anticipated Barriers, Events that Enabled Successful Community Entry, and Unanticipated Barriers.

Table 6.2

Barriers and Enablers to Community Entry

Barriers and Enablers to Community Entry	Critical Incident Case Number: Quote
Anticipated Barriers	
Conflicting Agricultural or Weather Cycles	Case 4: The evaluator stated that "when we go to the country … we tried [to determine] 'What are the agricultural cycles?'"
Conflicting Previous or Ongoing Research	Case 4: The evaluator stated that "when we go to the country ... to determine if … there aren't other duplicative types of research ... if there was other research in the community that could cause challenges … and what previous research has been conducted."
Problems, Rumors or Miscommunications	Case 4: The evaluator stated, "We started the formative evaluation in country one [1]…. We want to stay on top of any potential problems, rumors, or miscommunications."
Insufficient Resources	Case 4: The evaluator stated that "we go to the country … to determine if we have staff that can support the project ... to look at what was available, such as 'Is the Internet available in the community?'"
Securing Ministries' Support	Case 4: The evaluator stated that "there have been occasions where getting the health departments on board was challenging … the challenging part was that we needed to sit with them and explain how our proposal relates to their broader mission."

Critical Incident	
Barriers and Enablers to Community Entry	Case Number: Quote
Work, Residence, and Migration Patterns	Case 4: The team hired people to interview "the women in the population that we were studying. We wanted to understand what was occurring where they work, live, and travel .... what were especially busy times or when it's not convenient to talk to people or whether there are events that make it more difficult to gain access to people."
Unanticipated Barriers	
Civil Unrest	Case 4: "Unfortunately, the evaluation had to be put on hold due to civil unrest."
Ruling Against Human Subject Research	Case 4: "In the other country [country two (2)] ... the Ministry made a ruling that there was no basis to do human subject research in the country at all."
Enabling Events	
Unanticipated Enablers	
Circumcision's Public Health Role	Case 7: "The stigmatization against religion was not an issue in the field … because recipients of the training and the clients disassociated circumcision from its religious role to a public health role."
Political Support	Case 7: In 2009, the practice of male circumcision was supported by country's leaders and the department of health.

Successful community entry can "instill accountability and responsibility for what researchers learn to see … [and] maximize reciprocity for the construction and validation of instruments, findings, and conclusions by examining the multiple world views on the issue that collaborators provide" (RTI, 2004, p. 24). The category "Barriers and Enablers to Community Entry" occurred in two cases: Case Four and Case Seven. The following events occurred: Civil Unrest, Ruling against Human Subject Research, Lack of Resources, Conflicting Previous or

Ongoing Research, Conflicting Agricultural or Weather Cycles, Securing Ministries' Support, Threats to the Evaluation Process, Work, Residence and Migration Patterns, Circumcision's Public Health Role, and Political Support.

Manswell, Reid, and LaPoint (2004) indicated that understanding a program's cultural context involves understanding the "totality of the environment in which the program takes place" (p. 38). The environment includes the "geographic location, timing, political and social climate, and economic conditions" (Chourinard & Cousins, 2009, p. 479). In Case Four, the evaluation team took actions to pre-empt any possible barriers to community entry. Similar to Running Wolf, Soler, Manteuffel, Sondheimer, Santiago, and Erickson (2002), the expert evaluators in Case Four spent time in the community that was to be evaluated in order to mitigate problems, rumors, or misunderstandings, and become more familiar with the community's culture.

Bamberger, Rugh, Church, and Fort (2004) recommended that an evaluator "consider commissioning preparatory/exploratory studies prior to the arrival of time constrained external consultants" (p. 34). In addition, Bamberger et al. (2004) highlighted that seasonal cycles need to be considered when conducting an evaluation. In Case Four, the evaluation team conducted a resource analysis to determine the resources needed for conducting the evaluation, threats to the facilitators in the field, and threats to the evaluation process. To secure community entry, the evaluation team conducted preparatory studies to determine if there were agricultural, weather, work, migratory cycles, or residence patterns that could hinder their ability to conduct the evaluation.

Elkins (2010) defined peace-precarious situations as "characterized by recent, ongoing or recurrent periods of war and other violent disputes that abruptly or over time have significantly

destabilized but not obliterated core socioeconomic and political institutions" (p. 308). Peace-precarious situations can affect evaluation and program implementation. In Case Four, the evaluation team did not anticipate the outbreak of civil unrest in one of the countries being evaluated, and thus the team left the country due to security concerns. In another country site, the ministry made a ruling against human subject research, so the evaluation was placed on hold.

Several studies have been conducted on the difficulty of building trust between populations (Schwarz & Struhkamp, 2007). Dissonance in "power, status, and privilege" can be "particularly significant in diverse communities or communities with a history of exploitation and disempowerment" (Chouinard & Cousins, 2009, p. 480). The research conducted by Chourinard and Cousins' established that the evaluators often relied on persons or a group from the evaluated community to bridge the cultural gap between the evaluators and the stakeholders in the evaluation context. In Case Four, the evaluation team employed persons for "boundary spanning roles to provide a 'bridge' between external evaluators and the community to facilitate cultural and contextual understanding" (p. 480).

Bamberger, Rugh, Church, and Fort (2004) indicated that "it is important to consistently check for selection bias ... In some cases, non-participants may have been excluded or discouraged on the basis of their political affiliation (or lack thereof), sex, ethnicity or religion." In Case Seven, an enabling event occurred that assisted the evaluation's community entry. The first enabling factor was that the stigma associated with circumcision as a religious ritual was not an issue. Instead, the community accepted the intervention as a surgical intervention. The second enabling event was the lack of negative political interference (Bamberger et al., 2004). Furthermore, there was political support for the intervention. As a result, the evaluation did not encounter any barriers to community entry.

Barriers to Coordinating with Multiple Stakeholders

According to Mooney, "Co-ordination is an orderly arrangement of group efforts to provide unity of action in the pursuit of common goals" (1947, p. 5). Authors of the 1998 report, "The Programme Approach: Ownership, Partnership and Coordination" by the United Nations Development Programme's Evaluation Office, highlighted that among stakeholders "the understanding of what coordination actually means is not always clear" (http://web.undp.org/evaluation/documents/progapp.htm#fore, Section: Coordination Among Development Partners, para. 10). The authors also emphasized "the importance of partnerships and coordination" (Section: Forward, para. 3) in evaluation and as necessary "for the successful implementation" (Section: Forward, para. 6) of an evaluation.

A barrier to coordinating with stakeholders was narrated in Case Four. It was reported that the evaluation team considered the "process of coordinating multiple stakeholders an ongoing challenge … when conducting an evaluation in Sub-Saharan Africa." It was stated that

> a major evaluation challenge that we have encountered when conducting evaluations in Sub-Saharan Africa is that there are often multiple parties involved in an intervention …. That means we needed to coordinate with all of the parties involved in an intervention.

Barriers to Understanding the Evaluand's Context

According to Greene (2005), the evaluand context includes the demographic, geographical, cultural, and organizational elements present in the project or program being evaluated. Data analysis resulted in a category, Barriers to Understanding the Evaluand's Context. The category was further subdivided into Anticipated and Unanticipated Barriers to Understanding the Evaluand's Context. The barriers to understanding the evaluation context were indicated in four cases. The barriers were identified as Cross-Cultural and Language

Barriers, Two Countries and Their Populations, Logic Model and Misunderstood Program, and NGO and Culture (See Table 6.3).

Table 6.3

Barriers to Understanding the Evaluand's Context

Barriers to Understanding the Evaluand's Context	Critical Incident Case Number and Quote
	Anticipated Barriers
Cross-Cultural and Language Barriers	Case 7: The evaluator said, "We knew that the context did not transfer exactly from one evaluation site to another. We were aware that the messages needed to address the men's different ages, social class, and literacy rate, as well as their professional or laymen background."
Two Countries and Their Populations	Case 4: The evaluator stated that as part of the official protocol, "the team conducted structured observations to understand the evaluand context and the population being studied."
	Unanticipated
Logic Model and Misunderstood Program	Case 6: "Yet, no one seemed to be on the same page of what that logic model should look like. It was clear that there wasn't a clear understanding of the program …"
NGO and the Culture	Case 5: The evaluator stated that "the main challenge was trying to understand the contexts and trying to understand how the NGO was operating."

Understanding evaluation context is an important aspect of conducting an evaluation (Patton, 2008). Culture is "a cumulative body of learned and shared behavior, values, customs, and beliefs common to a particular group or society" (Frierson, Hood, & Hughes, 2002, p. 63). The research conducted by Kirkhart's (2010) stated that "culture is not a single unitary construct. Culture is multiply determined at both individual and collective levels. It represents an intersection of identifications held by individuals plus the identifications of collective grouping, such as organizations, institutions, communities, societies, and nations" (p. 402).

An evaluation team may struggle to understand the organization's [the NGO's] culture because "evaluation works within multiple collective and individual dimensions of culture" and because "[c]ulture is not fixed. Rather, the construct captures a fluid space in which the salience of each identification changes with the situation, role, task, and time" (Kirkhart, 2010, p. 402). Case Five indicated that an attempt to understand the evaluation context was an "ongoing challenge."

In Case Four, the expert evaluator stated that the evaluation team utilized structured observations and interviews to understand the evaluand context in two countries. The expert evaluator also stated that the use of structured observations was part of the authorized protocol. The evaluation team was unable to predict a civil war in one country and another country's decision to "halt human subject research." As a result, one evaluation was delayed until the former country was safe for the expert evaluators to return, and in the latter, the evaluation was put on hold until human subject research was allowed to continue.

Hood (2004) encouraged evaluators to conduct culturally responsive evaluations. Frierson, Hood, and Hughs (2002) stated that culturally responsive evaluators "honor the cultural context in which an evaluation takes place by bringing the needed shared life experience and understandings of the evaluation at hand" (p. 63). In Case Seven, the evaluation team was aware that one evaluation experience does not automatically transfer to another. Thus, the evaluation team modified the previously used evaluation instruments to the new evaluation context.

In Case Six, there was a lack of consensus on the program's theory of change as well as how the program was implemented. If a program's mission is not well defined and the means of achieving the program's mission are not clear, it is difficult to ascertain the reasons why a program's outcomes were or were not achieved (Stinchcomb, 2001). Further, it is difficult to

evaluate a program's expected outcomes if the means by which or the theory on how the program is to achieve its mission is not clear (Solomon, 2002). Thus, evaluators often develop logic models to understand the relationships between a program's inputs and outputs and to understand how the program is implemented (Bickman, Heflinger, Pion, & Behar, 1992; Julian, 1997; McLaughlin & Jordan, 1999; Weiss, 1997).

In Case Six, the evaluation technical support team assisted their evaluation partner to develop their program's logic model and establish a consensus of how the program would be implemented. The goals were to (1) assist stakeholders to articulate the program's theory of change, (2) explain the program's goals or outcomes (Herndanez, Hodges, Cascardi, 1998; McLaughlin & Jordan, 1999), and (3) provide a logical sequence of how the program's components relate to one another (Chen & Rossi, 1983, Hernandez, 2000; Rossi, Lipsey, & Freeman, 2004). Chen (1990) also indicated that the logic model could be utilized to compare the program's intended activities or goals to what is actually being implemented (Friedman, 1997; Mowbray, Holter, Teague, & Bybee, 2003).

Budget Constraints

Bamberger (2006) stated that budget constraints refer to "insufficient funds for the evaluation" (p. 51). Bamberger indicated that insufficient funds can occur when the "funds …were not included in the original project budget" (p. 23). In this study, three types of budgetary constraints were indicated: Bureaucracy and Non-Payment, Non-Negotiable Budget, and Insufficient Budget (See Table 6.4). Budgetary constraints were specifically mentioned in Cases Three, Five and Six.

Table 6.4

Budget Constraints

Budget Constraint	Critical Incident Case Number and Quote
Bureaucracy and Non-payment	Case 6: "... due to in-country administration and bureaucratic issues that are impeding release of additional funds from the U.S. government, the project is on hold."
Insufficient Budget	Case 3: "We needed a bigger budget so that we could have hired interviewers that had been properly trained and get more than self-reported risk behaviors … so we could increase our sample size."
Non-Negotiable Budget	Case 5: The evaluator stated, "in this evaluation there wasn't much room to negotiate the budget.

Bamberger (2006) indicated that the quality of the data collected and the sample size used in a study could be impacted by insufficient funds. In Case Three, the evaluation team utilized a small sample size due to an insufficient budget. In addition, the evaluation team was unable to collect biomarkers that could be used to determine if there was self-reported bias present in the evaluation.

The World Bank report, "Conducting Quality Impact Evaluations Under Budget, Time, and Data Constraints," (2006) recommended that if an "evaluation cannot be conducted within these constraints [time, budget, data availability], the resources and time frame must be renegotiated, the scope and objectives of the evaluation revised, or the evaluation cancelled" (p. 26). In Case Five, the ability to gather data was hampered by the inability to negotiate a larger evaluation budget. The possibility of bias in the data was a concern; therefore, a new evaluation was recommended to "take another look at the whole thing." In Case Six, the evaluation was put on hold because funds were not released to the evaluation team's partner.

Evaluation Resistance

According to Taut and Brauns (2003), "Programme evaluators frequently encounter resistance from individuals affected by evaluation" (p. 247). Table 6.5 lists the types of evaluation resistance the expert evaluators encountered: Interview Avoidance, Evaluation Avoidance, and Resistance to Evaluation Findings and Report.

Table 6.5

Evaluation Resistance

	Critical Incident
Evaluation Resistance	Case Number and Quote
Interview Avoidance	Case 1: The evaluator concluded that the "interviewee was avoiding being evaluated."
Evaluation Avoidance	Case 2: "I [the evaluator] know it was my assumption, but I think they were avoiding getting evaluated ..."
Resistance to Evaluation Findings and Report	Case 2: The evaluator said, "I had evidence that supported my findings. I tried to explain it to them [that there was no impact] but they just didn't seem to get it."

Taut and Brauns (2003) stated that "[r]esistance occurs throughout the evaluation process, from the inception of an evaluation to the utilization of its findings" (p. 247). The research conducted by Schwant and Dahler-Larsen (2006) established that "evaluation resistance can take many forms or have many different motivations" (p. 406). Evaluation resistance may be attributed to such factors as a stakeholder's attitude toward evaluation (Patton, 1997), a stakeholder's need to preserve the status quo (Weiss, 1998), the power related interactions between stakeholders (Bonoma, 1977), or a stakeholder's perception of supposed encroachment on his or her autonomy (Wottawa & Thierau, 1998).

It was observed that evaluation resistance occurred in at least two cases. In Case One, a department official was resisting being interviewed. In Case Two, the lead consortium did not schedule the necessary evaluation interviews and focus groups in an attempt to prevent the evaluation's implementation. Evaluation resistance occurred in Case Two when the evaluation sponsors rejected the evaluation findings and report.

Successful Collaboration with Stakeholders

Mattessich, Murray-Close, and Monsey (2001) defined collaboration as "a mutually beneficial and well-defined relationship entered [into] by two or more organizations to achieve common goals" (p. 39) that requires comprehensive planning and communication on many levels. According to Gajda (2004),

> an increasing number of organizations are coming together to address complex societal issues. Most intentional, inter-organizational collaboratives (i.e., strategic alliances) articulate the collaborative effort as the primary method for achieving ideal short and/or long-term goals that would not otherwise be attainable as entities working independently. (p. 65)

Collaboration with stakeholders. The Case Seven quote, given below, states that the evaluation team worked collaboratively. Collaboration among the evaluation team resulted in the following: they (1) were able to address the cultural challenges that the team encountered, (2) developed evaluation instruments such as interview guides and surveys, and (3) determined when to conduct the surveys.

> We worked very collaboratively … everyone in the team brought in their own experience and expertise to better massage the cultural challenges. … everybody was contributing to the development of the instruments and to incorporate the specific cross-cultural issues …. We further collaborated as we determined when to ask the young men specific [evaluation survey and interview] questions.

Threats to High Quality Data

The characteristics of high quality data are accuracy, completeness, reliability, timeliness, confidentiality, precision, integrity, and measurability (MEASURE Evaluation, 2007; Salabarría-Peña, Apt, & Walsh, 2007). In this study, three types of data constraints were identified: Lack of National Guidelines, Lack of Research Expertise, and the Need to Track Trends (See Table 6.6).

Table 6.6

Threats to High Quality Data

Threats to High Quality Data	Critical Incident Case Number and Quote
Lack of National Guidelines	Case 6: "We found that national guidelines on the supervision of home-based healthcare did not exist …. The evaluator asked, "How can the quality of supervision of home-based healthcare be evaluated if there are no standards to compare it to?"
Lack of Research Expertise	Case 3: The evaluator stated that "the NGO was good at implementing the intervention, but they just didn't have the research expertise necessary to gather data that was of good quality …. It turned out to be a huge disaster."
Need to Track Trends	Case 4: The evaluator reported that the evaluation "utilized tools such as a GIS system to measure coordinates and identify specific trends."

The research conducted by Bamberger et al. (2006) established that evaluators often work under data constraints or requirements. It was observed that the data quality concerns were shared in Cases Three, Four, and Six. In Case Three, the data gathered was of poor quality. In Case Four, the evaluation team adapted their data collection methods to determine trends in the data. In Case Six, there was a lack of measurable indicators for comparing the program outcomes.

Threats to the Evaluation Design

Evaluation "designs specify the organization or structure for collecting data" (Fitzpatrick et al., 2011, p. 343). Data analysis resulted in a category, Threats to the Evaluation Design. Table 6.7 lists the five threats to the evaluation design: Inability to Provide Input to the Evaluation Design, NGO's Credibility, Non-Communicative Evaluation Funder, Bureaucracy and Miscommunication, and Unsatisfactory Protocol.

Table 6.7

Threats to the Evaluation Design

	Critical Incident
Threat to the Evaluation Design	Case Number and Quote
Bureaucracy and Miscommunication	Case 3: "there was a bureaucratic mix-up … [and] we never got the reviewers' comments … Headquarters … shut down the comparison portion of the study.
Inability to Provide Input to the Evaluation Design	Case 5: The evaluator stated that "the agency had set up the evaluation before I arrived … [and] did not give me the opportunity to provide any meaningful input in the design of the evaluation."
NGO's Credibility	Case 3: The evaluator said that "because the intervention had started and the NGO had already chosen the micro-credit beneficiaries, we couldn't do a randomization evaluation … going back on their [NGO's] promises would really damage their credibility in the community."
Non-Communicative Evaluation Funder	Case 2: "I [the evaluator] did try to make [funder engagement] happen."
Unsatisfactory Protocol	Case 6: "Once we took a look at the protocol … we realized that it was of low quality …"

During evaluation design, an "appropriate design should be selected and its implications discussed with stakeholders to reduce the threats to an evaluation design's reliability, validity, and utility" (Fitzpatrick et al., 2011, p. 343). In addition, an expert evaluator's credibility and

integrity are also at risk when they present their own best judgments concerning an appropriate evaluation design to stakeholders as well as when they ask stakeholders to trust the evaluator's judgments (Patton, 2008). In this study, threats to the evaluation design occurred in four cases.

In Case Three, the evaluator was inexperienced in navigating the funding agency's bureaucracy. In addition, there was a miscommunication regarding the evaluation approval process. Consequently, the initial evaluation design was implemented before it was authorized. When the funding agency learned of the unauthorized evaluation, the evaluation design was significantly modified with only a portion of the original design allowed to continue.

Problems were found to occur before the evaluation team submitted a design to their main office for approval. The evaluation team had originally hoped to conduct an evaluation with a randomized evaluation design. However, they were unable to randomize who would participate in the evaluation as the NGO working with the evaluation team had already begun the intervention and selected the participants for the program. Thus, a change in who would participate in the program so that the evaluation team could conduct a randomized study would ruin the NGO's credibility.

In Case Five, the evaluator was not given the opportunity to provide input into the evaluation design. The expert evaluator had completed the initial evaluation but was concerned about potential bias present in a "precooked" evaluation. Therefore, the expert evaluator recommended that a new evaluation be commissioned.

In Case Two, the evaluation design was threatened when the agency funding the evaluation did not engage in the evaluation design process. As a result, the organization that commissioned the evaluation was not involved in the evaluation's design. The expert evaluator

relied on her own judgment and training to determine that an exploratory approach was necessary to conduct the evaluation.

In Case Six, a poorly designed evaluation protocol was submitted to the evaluation team providing technical support. The evaluation team decided to provide their evaluation partner with training on how to design the evaluation.

Time Constraints

Time constraints refer to not having sufficient time to evaluate a program (Bamberger, Rugh, and Mabry, 2006). Data analysis of the expert evaluators' narratives resulted in two subcategories regarding Time Constraints. The subcategories were: Insufficient Time and Sufficient Time (See Table 6.8). The constraints were as follows: Bureaucracy and Non-Compliance, Rapid Data Gathering, Evaluation Commissioned at End of Project's Life Cycle, Hurried Evaluation, and Logistics and Methodology.

Table 6.8

Time Constraints

	Critical Incident
Time Constraints	Case Number and Quote
Insufficient Time	
Bureaucracy and Non-Compliance	Case 4: The evaluator "had to defund the evaluation because the award recipient was not doing what they were supposed to do … [it] added six months to the implementation timeline."
Rapid Data Gathering	Case 4: The evaluator stated, "We needed to gather the data quickly."
Evaluation Commissioned At End of Project's Life Cycle	Case 5: The evaluator stated that "there wasn't very much time … The end of the project was coming up, so it was close to the wire."

	Critical Incident
Time Constraints	Case Number and Quote
Hurried Evaluation	The evaluator stated that, "only three months [were] allocated to conduct the evaluation … It was much more hurried.
Sufficient Time	
Logistics and Methodology	Case 7: the evaluation team was "not pressed with time issues … in terms of logistics … and methodology"

In Case Four, time constraints occurred due to the bureaucratic bottlenecks caused by the lengthy process of approving a protocol and funding an evaluation. Bottlenecks occurred because the staff at the funding organization's headquarters was overworked, and there was insufficient staff to expedite the approval of the evaluation protocol. This is consistent with the research conducted by House, Haug, and Norris' (1996) on evaluation and bureaucracy. House et al. stated that "core evaluation work performed by evaluation personnel consisted of many tasks … [where the evaluation staff] saw themselves as 'hopelessly understaffed'" (p. 138).

The research conducted by House et al. (1996) established that "[while] evaluation staff design and manage the evaluation they [also] contract out the day-to-day work. Contracting out evaluations, which is a necessity given the 'down-sizing' of government, means that the evaluators are engaged in monitoring and negotiating details among program staff and contracts" (p. 136). The research conducted by Bamberger et al. (2006) highlighted that time pressures can occur because an outside consultant "has traveled a long distance so the timing of the interventions [or evaluation] has to be well planned or coordinated" (p. 81). In Case Four, the evaluation contract was de-funded because of a contractor's non-performance. As a result, the team had to re-advertise the contract. The need to find a new contractor added six months to the evaluation timeline.

Bamberger et al. (2006) stated that

> Given the need to adapt many of these methods to the time constraints under which most program evaluations take place, international development literature (which is apparently not well known in the United States) has developed where the focus is on the rapid methods of data collection. (p. 76)

In Case Four, the evaluation team needed to gather data quickly. Thus, the team had to use rapid assessment procedures. The research conducted by Bamberger (2000), Beeve (2001), Handwerker (2001), and Scrimshaw and Hurtado (1987) provides guidelines and/or case studies on the use of rapid assessment procedures.

In Case Five, the data gathering approach was influenced by time constraints because the evaluation was commissioned late in the project's life cycle. The experience in Case Five is consistent with Bamberger et al.'s (2006) statement that time constraints often occur "when the evaluator is not called in until the project is already well advanced and the evaluation has to be conducted within a much shorter period of time than the evaluator considers necessary" (p. 23). The evaluator stated that due to insufficient time, he did not have time to conduct a literature review. He also stated that he needed to hurry when conducting the evaluation.

In Case Seven, having sufficient time assisted in implementing the evaluation plan and methodology. It was stated that the evaluation team was "not pressed with time issues … in terms of logistics … and methodology …" This statement may indicate that time constraints can have a negative or positive influence in implementing the evaluation plan and methodology.

Critical Incidents Summary Table

Table 6.9 provides the highest level of abstraction of the critical incidents encountered. The first column in the table lists the critical incidents found in the expert evaluators' narratives. The second to the eighth column lists this study's case numbers. The rows in the table list a specific critical incident and the case numbers indicate the specific case for the critical incident.

Table 6.9

Critical Incidents

Critical Incidents	Case Number						
	1	2	3	4	5	6	7
Barriers to Communicating with Stakeholders				x		x	x
Barriers and Enablers to Community Entry				x			x
Barriers to Coordinating with Multiple Stakeholders				x			
Barriers to Understanding the Evaluand's Context				x	x	x	x
Budget Constraints			x		x	x	
Evaluation Resistance	x	x					
Successful Collaboration with Stakeholders							x
Threats to High Quality Data			x	x		x	
Threats to the Evaluation Design		x	x		x	x	
Time Constraints (Insufficient or Sufficient Time)		x	x	x	x	x	x

Research Question Two:

What factors influenced expert evaluators' naturalistic decision making during evaluations of HIV/AIDS health education programs?

This section provides the results of the data analysis, which led to the categorization, and identification of the types of factors that influenced expert evaluators' decision making during critical incidents. Tables and narratives have been used to report the results. The narratives provide a definition of the categories and a brief description of the types of critical incidents. Tables were used to report the results when more than one factor type was reported. The tables generated during data analysis were very lengthy; therefore, they were aggregated. In addition, a maximum of two examples were provided to depict the types of factors reported by the expert evaluators in their narratives.

Two kinds of tables were created. One table lists the types of factors that influenced the expert evaluators' decision making and their matching case numbers and quotes. It provides a

high level description of the critical incidents reported by the expert evaluators. The other table, with the highest level of abstraction, provides a high level overview that lists the factors and the case numbers for cases in which the factors occurred.

In 1970, Patton studied 20 federal health evaluations to determine the factors that influenced evaluation use. Patton's research identified a factor labeled as the Personal Factor. Alkin (1985) developed an evaluator's decision making guide to enhance evaluation use. In the guide, Alkin identified three types of factors that influence evaluation use, which are as follows: Context Factors, Evaluation Factors, and Human Factors.

It was found that personal and human issues were so intertwined that it was impossible to separate them into individual categories. Therefore, they were combined under the category Human Factors. Data analysis resulted in a new category, Real-World Decision Making Factors. The following categories emerged: Context Factors, Evaluation Factors, Human Factors, and Real-World Decision Making Factors.

Context Factors

According to Alkin (1985) context factors

> refer to the specific setting in which the evaluation is conducted. Included here are such elements such as, fiscal constraints on the evaluation, the length of time of the project operation, and the social/political climate surrounding the project [evaluand]. (p. 24)

The type of context factors that influenced the evaluators' decision making were: Civil Unrest, Country's Agricultural and Weather Patterns, Country's Infrastructure, Culture, Evaluation Participants' Work, Residence and Migration Patterns, Geographical Barriers, Language, Politics, and Power.

Civil unrest. In Case Four, civil unrest occurred in one of the countries in which the evaluation team was conducting structured interviews. In Case Four it was stated that "In one of

the countries [where the evaluation team was working], civil unrest broke out," so the evaluation team decided to "pull out." Due to the civil unrest, the evaluation team made the decision to place the evaluation "on hold" until it was safe for the evaluation team to return.

Culture. In Case Seven, it was reported that "… the clients' and the medical personnel's cross-cultural, cross-religious, cross-racial, and socio-economic status needed to be addressed in the research instruments and the protocol." Case Seven provides multiple examples of the challenges encountered when trying to be culturally sensitive. Decisions were made to address culture and language issues.

Case Four encountered an instance in which "the ministry decided that they would not allow any human subjects research." Due to the ministry's decision, the evaluation team decided to put the evaluation on hold. In Case Four, cultural issues were addressed by employing interviewers from the community in which the evaluation was to be implemented. The expert evaluator stated, "I think it's important that people can relate to the interviewers rather than having a white person from the United States coming in to conduct the interview."

Country's agricultural and weather patterns. In Case Four, during structured observations the evaluation team attempted to determine "agricultural cycles." The evaluation team also "looked for cycles [such as seasons or weather patterns] that we need to be aware of." Thus, data was gathered to determine when to implement the evaluation.

Country's infrastructure. In Case Four, the evaluation team decided to determine "what was available [by asking questions] such as, 'Is the Internet available in the community?'" Resource assessment was used to identify the possible barriers to community entry.

Evaluation participants' work, residence, and migration patterns. In Case Four, it was reported that the team hired consultants to interview "the women in the population that we

[the evaluation team] were studying." It was decided to conduct the interviews because the evaluation team "wanted to understand what was occurring where they [the evaluation participants] work, live, and travel." The goal was to determine when it would be "convenient" to conduct the evaluation interviews and if "there are events that make it more difficult to gain access to people [evaluation participants]."

Geographical barriers. In Case Seven, the evaluation team "communicated through email and met via the Internet. It made life easier to use the technology at our disposal. It was a way to bridge over geographical gaps." The use of technology assisted in keeping the channels of communication open as the evaluation team members were geographically dispersed. Technology was a tool used to overcome the geographical challenges that the team encountered.

Language barriers. In Case Four, "The evaluation team hired interviewers and observers who could speak the local languages of the two countries and were from the surrounding community." It was anticipated that there would be language barriers because different languages were spoken in the communities being evaluated. The team was concerned that the language barriers would affect data collection. The expert evaluator stated that the evaluation team decided to hire interviewers who spoke the evaluand communities' local languages.

In Case Seven, the evaluation team modified the data gathering instruments so that "90% of the text was in English and 10% was in the local language." The modifications were made because the evaluation team observed that:

> the male nurses were sometimes explaining the materials that were originally in English in their local language. We [the evaluation team] were curious as to why they needed to do that because generally the population spoke and understood English. The nurses explained to us that some words do not exist in their local language. They also explained that some words have a more powerful meaning in

> their local language, so they explained the concepts written in English in their language.

Decisions were made to address language barriers that could limit the evaluation participants' comprehension of the data gathering instruments.

Politics. In Case One, it was stated that "HIV is politicized" and that the government and the health department held the political view that HIV does not exist or that HIV does not cause AIDS. The reaction to the politics surrounding HIV and AIDS influenced the evaluator's decision to confront the interviewee who avoided being interviewed. It was reported that "[as an evaluator] you become a player in all of that drama. It was a running challenge throughout the evaluation."

In Case Five, it was indicated that the research design might have been planned in a way to "[p]rotect the project. Protect the agency. Protect the Ministry of Education." It was reported that "often the evaluation just becomes a political exercise" and that politics has an influence on the evaluator's decision making.

Power. In Case One, the evaluator indicated that power has an influence over an expert evaluator's ability to "navigate the power differences [between evaluator and evaluation participant]." A previous evaluation experience guided the evaluator's response to the "power differences." In addition, the evaluator also indicated that the authority and years of research on HIV and AIDS influenced the evaluators' decision to confront the hostile department official. Nevertheless, the evaluator was concerned about not abusing the "authority that comes with my role as an evaluator."

In Case Six, it was reported that there was a power struggle among the evaluation partners that needed to be addressed; the power struggle affected the interactions among the evaluation partner's PI, Co-PIs, and the evaluation technical support team. Due to the power

struggle, it was decided “to provide training on leadership and management training” as a way “to address the issues regarding local team member participation and empowerment.” After the training, the training recipients reported, “How life-changing this process [the training] was for them ... They used the word ‘empowerment’ as a way to explain … what they got out of the experience.”

Table 6.10 depicts the types of contextual factors identified as a result of the data analysis. The first column in the table lists the contextual factors found in the expert evaluator’s narratives. The second to the eighth column indicate the numbers of cases in which the specific type of factors can be found in the narratives. The rows indicate a specific type of context factor and its matching case numbers.

Table 6.10

Context Factors

	Case Numbers						
Context Factor	1	2	3	4	5	6	7
Civil Unrest				x			
Country's Infrastructure				x			
Country's Weather Patterns				x			
Country's Agricultural Patterns				x			
Culture	x	x	x	x	x	x	X
Evaluation Participants’ Work, Residence and Migration Patterns				x			
Geographical Barriers							x
Language				x			x
Politics	x		x	x	x	x	x
Power	x	x			x	x	x

Evaluation Factors

According to Alkin (1985, p. 24), an evaluation factor “involves the evaluation itself. Included here are such elements as the evaluation procedures used, the kinds of information

collected, and the ways in which the information is reported." The results indicated that the types of evaluation factors present in the expert evaluators' narratives were as follows: Bias, Budget, Data Quality, Evaluation Usability, Implementation Barriers, Consent, Confidentiality, and Privacy, Lack of Research Skills, Life Cycle, and Poor Planning.

Bias. In Cases Three and Five, bias was a factor that influenced the expert evaluator's decision making. In Case Three, "bias [was] a concern." It was explained that "reported risk behaviors tend to go down just because of the introduction of the intervention. So you [the evaluator] wonder[ed] if the changes are true or not." It was reported that "hiring the implementing partner, an NGO, to conduct the evaluation's data gathering introduced bias into the evaluation." The expert rationale was that it "was not good to have the NGO to do their evaluations because they don't have the research expertise … [and] NGOs are trying to get money for their projects." It was pointed out that "just from hearing people talk, it is good practice to have an external evaluator … [because] there is a stronger potential for bias [when an NGO is conducting an internal evaluation]."

In Case Five, there was concern about the "pre-cooked" evaluation design and the fact that the evaluator was an internal evaluator. It was narrated that:

> In a precooked evaluation, you worry about bias because there was a bias built into it. It was not truly independent because I was a member of the funding agency. I did come in with a critical frame of mind. Looking back, it was by no means as critical as I have been as an independent evaluator. You have a different stance when you are internal evaluator. You tend to go with the flow. You work within the architecture and you work within the assumptions of the organization. You may identify problems, but you work within the framework of the organization.

Budget constraints. In Case Three, budgetary constraints influenced the evaluation team's decision to "hire the NGO to conduct the data gathering." It was reported that the

evaluation team "couldn't afford to hire a high-powered research group, [so] what we did was ask the NGO to do the evaluation."

Due to a limited budget, the team was unable to gather biological markers that would have provided comparison data to determine if there was self-reported bias. In addition, a larger budget would have provided the team an opportunity to increase the evaluation sample size. It was stated that

> a bigger budget [was needed] so that we [the evaluation team] could have hired interviewers that had been properly trained and get more than self-reported risk behaviors … like some biological markers, like STIs (Sexually Transmitted Infections). We also needed a bigger budget so we could increase our sample size. We had a pretty small sample size of about 700-800 people.

Due to insufficient funds, it was "decided to recommend the commission of a new evaluation." In Case Five, "the evaluation team did not have enough funds, so they paid for the evaluation expenses through the evaluation funding organization's budget." The evaluation team "was part of the funding organization, so we just managed it with our resources."

Further, it was also necessary to "find ways … [to] bring in other people" but "again, that required extra budgeting." As a result of the insufficient budget, "a short evaluation report" was generated. The team did not present a "formal evaluation presentation ... [because a formal presentation] required another visit, which adds to the time and budget of an evaluation." Eventually, it was "recommended that a more detailed evaluation should take place … because it required a fresh look at it all."

Consent, confidentiality, and privacy. When conducting research it is often necessary to obtain informed consent from evaluation participants (Walker, Hoggart, & Hamilton, 2008). In Case One, the requirements for consent for participation in the evaluation were not as strict as the requirements for participation in the research. It was stated that the department official's

"consent to participate in the evaluation was required," which encouraged continued attempts to secure the official's participation in the interview process. The distinction between the evaluation and the research was explained with the following statement, "Evaluation is different from a research project. If this was my research project and she [the HIV health department official] was reluctant to talk to me, I wouldn't have any business pursuing it further."

In Case Four, the organization sponsoring the evaluation required the evaluation participants to be educated about the research participation consent. The participation consent was not only a requirement, but also provided maximum privacy and ensured data confidentiality. In Case Four, the team "tried to have maximum privacy when conducting a study so that we can assure both participants' and data confidentiality." It was stated that

> protecting participants' confidentiality is one of our main responsibilities, so we educate participants about the consent process. It is also critical that we let them know why the research is important. Their increased understanding about the consent process and about research or evaluations can benefit their community and inform the design of future health programs.

The evaluation team considered "engaging and involving persons in the research" as "a part of the education process."

Criteria. In all cases, criteria were used to determine the evaluand's merit. In Case Two, for example, decisions were guided by efforts to determine the program's impact. The evaluator reported the findings, although she encountered resistance. The evaluator reported "results indicating that it was really either not being the way to go [to use the consortium model] or it was not being implemented correctly, and because of that, it was impossible for me, in that context, to find out about any kind of impact."

Data quality. In Case Three, the information gathered from a literature review influenced the evaluation decisions, such as to which evaluation questions should be asked. The

evaluator indicated that “A literature review … found that national guidelines on the supervision of home-based healthcare did not exist.” “Based on our [this] finding, we decided to ask the evaluation questions, ‘How is the supervision of home healthcare being implemented?’ [and] ‘Is supervision being conducted? And if so, how?’”

Case Four depicts an event in which poor quality data was gathered. It was reported that due to a small budget, the evaluation team “hired the NGO that was implementing the program to gather the evaluation data.” It was found that “the NGO was good at implementing the intervention, but they just didn’t have the research expertise necessary to gather [survey] data that were of good quality.” Furthermore, the statisticians hired to conduct the data analysis said that “they couldn't link the baseline to the follow-up information … [nor] could they link the sexual behaviors data to the economic loan data.” This event was considered “a huge disaster.” [The evaluators] “took a look at the sexual behavior baseline data, and there was a 40% non-response to some of the behavioral questions.” They decided, “You can't really use variables like that.” Compounding the problem was that “for the micro-credit part of the intervention, the NGO had these extensive economic interviews and transactional behavior interviews … [but the statisticians] weren't able to link [the data to] any changes in behavior.” As a result, the data from the survey was not used. The qualitative data gathered from the interviews was of “reasonable quality,” which was used.

Evaluation utilization. Efforts were made in all the narratives to ascertain whether the evaluation process was useful or that the evaluation findings were utilized. In Case Two, it was considered a “matter of practice to use the evaluation process for evaluation capacity building” by “explain[ing] what evaluation is, how it should be working” and how to “do it.” This approach was influenced by Patton’s Utilization Focused Principles.

Case Six referred to Patton's Utilization Focused Principles. It was stated that the employer's "mandate is to increase evaluation capacity for program evaluation" and that the evaluation technical support team is "concerned about stakeholder use or ensuring that evaluations are feasible and utilized." Thus, the evaluation team focused "on the lack of supervision standards in supervision of home-based care … [because] the evaluation had the potential to inform home healthcare standards."

Bureaucracy. In Case Three, it was observed that challenges occurred due to miscommunication and a "bureaucratic mix-up." The evaluation team submitted a conceptual proposal of the evaluation to their organization's headquarters, and the team "thought that the evaluation proposal had been approved [so we started implementing the evaluation]. The evaluator found out that a "bureaucratic mix-up," had occurred. The evaluator stated that when "headquarters found out about it [that we had begun implementing the evaluation] they said, 'How could your study have started without our approval of the protocol?'" Due to this event, headquarters "shut down the comparison study portion of the quantitative study."

Civil unrest. In Case Six, the evaluation team encountered "civil unrest" in one country and resistance to "human subject research" in another. As a result of these two events, the evaluation team decided to "pursue alternative study sites." It was considered an "unexpected barrier to community entry."

Lack of research skills. According to Kniker (2011), lack of "expertise can be a problem" and evaluators "need to master social science research methodology" (p. 68). In Case Three, the evaluation team hired the program implementation partner to assist the team in conducting the evaluation. The NGO "indicated that they had the necessary research skills" to assist in gathering the evaluation data. The data collected indicated that the NGO was not

equipped to assist in the data collection process. Due to the NGO's lack of research skills, the data obtained was of "poor quality." It was observed that the data was of such poor quality that the "statisticians hired to conduct the data analysis could not link the baseline data with the data the NGO gathered."

Life cycle. Morrell (2010) "takes the perspective that both innovations and evaluations go through life cycles, and that interactions between evaluation and innovation depend on where each is in the life cycle" (p. 10). In all of the cases, the decision to conduct a formative or summative evaluation was influenced by the evaluand's phase in its lifecycle. For instance, in Case Four, the evaluator decided to conduct a formative evaluation to assist in the "design of the intervention's educational component." In Case Five, it was decided to conduct a summative evaluation as the evaluation funders were evaluating an intervention at the end of the program's lifecycle.

Poor planning. In Case One, the research participant explained that the evaluation participants could be resistant to being evaluated. This was attributed to "poor planning from higher up [donors and program implementers]," resulting in participants that are "over-evaluated" and "over-researched." He attributed the over-evaluation of programs to "the lack of coordination among program funding agencies." Consequently, program managers are overwhelmed by the amount of reporting to be completed in order to comply with "onerous project requirements with many lines of reporting." This causes participants to be "trapped in a duplication of a lot of effort" where the "effort can increase exponentially based on how many funders you have and who they are." Thus, as a result of poor planning, the evaluation environment can become hostile.

A table was generated to depict the types of evaluation factors identified from the data analysis (See Table 6.11). The first column in the table lists the types of evaluation factors found in the expert evaluators' narratives. The second to the eighth column indicate the case numbers in which the factor types were found in the narratives. The rows indicate a specific type of evaluation factor and its matching case number(s).

Table 6.11

Evaluation Factors

Evaluation Factor	Case Number						
	1	2	3	4	5	6	7
Bias			x		x		
Budget Constraints		x	x		x	x	
Consent, Confidentiality, Privacy	x			x			
Criteria	x	x	x	x	x	x	x
Data Constraints	x	x	x	x	x	x	x
Evaluation Utilization	x	x	x	x	x	x	x
Lack of Research Skills			x				
Life Cycle	x	x	x	x	x	x	x
Poor Planning	x						

Human Factors

In 1985, Alkin published a book that studied the factors influencing evaluation use. The research conducted by Alkin (1985), stated that human factors "includes user and evaluator characteristics that might have an effect on use" (p. 24). Alkin further divided the "Human Factors" into two sub-categories: Evaluator Characteristics and User Characteristics. The Evaluator Characteristics were further sub-divided into the following: Commitment to Use, Willingness to Involve Users, Choice of Role, Rapport with Users, Credibility, Background, and

Identity. The User Characteristics were further sub-divided into the following: Identity, Interest About the Evaluation, Commitment to Use, Professional Style, and Information Processing.

The category Human Factors was modified to include the user and evaluator characteristics, which affected the evaluator's naturalistic decision making when evaluating an HIV/AIDS health education program. The category Evaluators' Human Factors was modified to indicate the Evaluator's Attitude Toward the Evaluation, Evaluator's Attitude Toward Stakeholders, Evaluator's Attitude Toward the Evaluand, and the Evaluator's Role. A category labeled Evaluation Team Members' Attitudes Toward One Another" was added, as the evaluators were often working with an evaluation team to conduct the evaluation. The category Stakeholder's Human Factors was modified to indicate the Stakeholders' Attitudes Toward One Another, Stakeholder's Attitude Toward Being Evaluated, Stakeholder's Attitude Toward the Evaluator, Stakeholder's Attitude Toward the Evaluand, and the Stakeholder's Role.

Patton (2008) stated the following: "… personal factors are the presence of an identifiable individual or group of people who personally care about the evaluation and the findings it generates" (p. 69). In this study, the personal factors — defined by Patton as the "stakeholder's interest in the evaluation" (p. 69) and by Weiss (1990) as "a person's interest, commitment or enthusiasm" (p. 177) — were interpreted as the evaluators and stakeholder's attitudes toward the evaluation. As a result, personal factors were incorporated into the human factors subcategories labeled Evaluator's Attitude Toward the Evaluation and Stakeholder's Attitude Toward the Evaluation.

Evaluation team members' attitudes toward each another. In Case Six, efforts were made create a respectful and open work environment. For example, it was stated that

> I try to create an environment where there is respect. I don't create an environment that is hierarchical or where *I* know it all. I let people know that I

> am here to work with you. I create a space so that whatever we need to get out of the way — well, let's get it out of the way so you can be real with me.

In Case Seven, it was reported that "There weren't any dramatic issues." It was observed that when referring to the evaluation team's collaboration, the evaluation team was "on the same page."

Evaluator's attitude toward stakeholders. Case Two provided examples of the evaluator's attitudes toward stakeholders. The attitudinal factors were revealed through the evaluator's point of view toward the evaluation sponsor as well as the lead consortium. For example, there was frustration when the organization sponsor did not engage in a discussion that would help the evaluator understand the goals of the evaluation, determine the research questions, and decide whom to interview. The evaluator stated, "I doubt that it would have helped [to continue to try to engage with the funder] considering the situation." The frustration with the lead consortium was expressed when the evaluator said, "They knew darn well that I was coming …"

Evaluator's attitude toward the evaluand. Case One expressed the evaluator's attitude toward the evaluand. He stated that the program was a "mess" and was not meeting the program participants' needs. Unlike Case One, there was a positive attitude toward the program being evaluated in Case Seven. For example, it was stated in this manner:

> I would like to share with you an intervention currently being evaluated that relates to an exciting strategy that currently assists in the reduction of HIV transmission from infected females to their heterosexual male partners." The evaluator's tone indicated that the expert evaluator excited that the procedure could "reduce the transmission of the HIV virus from infected females to uninfected males by 65%.

Evaluator's roles. Volkov's (2011) research found that evaluators could take on multiple roles when conducting an evaluation. Volkov defines role as "an explicitly and

implicitly expected function performed and behavior associated with a particular position in an organization" (p. 27).

Evaluators are often categorized as internal or external evaluators. In Case One, the research participant indicated that his role as an external evaluator provided him with the necessary authority and power. He said, "You know, we [external evaluators] have a lot of power as outside people." His decision to confront the hostile interviewee was influenced by the power he enjoyed as an external evaluator.

In Case Five, the research participant indicated that as an internal evaluator he did not have the power he needed to negotiate with the organization that funded the evaluation. He stated that as an internal evaluator, "[y]ou work within the architecture and you work within the assumptions of the organization … [On the other hand,] as an external evaluator, you have much more power and voice."

In Cases Two and Six, their evaluation roles influenced the evaluators' decision making. In Case Two, the evaluator considered herself as a "development worker" and "teacher." In that role, her job was to "conduct evaluations for the betterment of the people in developing countries" and focus on "evaluation capacity building." In Case Six, the research participant considered herself as a teacher; this perception influenced how technical assistance was delivered.

Stakeholders' attitudes toward one another. In Case Six, there was conflict among the stakeholders, as well as tension among the evaluation partner's Co-PIs and the PI. The evaluation team decided to add a leadership training component to their technical support workbook. In Case Two, the evaluator stated that the stakeholders were frustrated with the lead consortium and program funding agency because of the poorly dispensed program funds.

Stakeholder's attitude toward being evaluated. In Case One, interviewee resistance was occurred because as the evaluator reported, as the official being interviewed considered the evaluation "an inconvenience and a waste of her valuable time." The official considered the evaluation "redundant" because she had been "reviewed six months ago."

In Case Six, stakeholders — the evaluation partner — requested technical support from the evaluation team. The initial response from the evaluation partner was "lively" with "push back." However, when the situation changed, the evaluation partner's Co-PIs "didn't feel like they could speak; the PI (Primary Investigator) was speaking on behalf of everybody."

Stakeholder's attitude toward the evaluator. In Case One, the evaluator reported that the government health department official considered the evaluator to be a low-level consultant. Thus, the official's attitude toward the evaluator was reflected in her behavior when she refused to participate in the interview process. In Case Two, the research participant reported that the organization funding the evaluation also expressed anger. She stated, "They were quite angry and said, 'This is not an impact evaluation. This is a site visit report.'" Thus, the evaluator reported that "They never hired me again."

Stakeholder's attitude toward the program. The evaluator reported that project participants expressed frustration because the money they needed "wasn't coming through." As a result, "People [the evaluation participants] were very irritated about the whole program, the [lack of] funding, and so on."

In Case Seven, the evaluator stated that the stakeholders endorsed the program's initiative and evaluation. The intervention was viewed as a "health benefit," thereby indicating support for the program. For example, the evaluator stated that

> While there is a stigma associated with HIV/AIDS, I [the evaluator] want to highlight that without our needing to intervene; there was a detachment of

circumcision and circumcision as a ritual performed in Islam and Judaism. The clients and medical professionals viewed circumcision as a health benefit.

Stakeholder's role. The evaluator's decision making was influenced by the stakeholders' role. It was observed that the stakeholder's role in the narratives was diverse. At times, the stakeholders were the evaluation participants, clients, and members of the evaluation team, an agency funding the evaluation, or a recipient of program services. For example, in Case One, the stakeholder was confronted as she was considered accountable to the government, taxpayers, program funders, and program recipients. In Case Six, the stakeholders were clients and members of the evaluation partnership. The evaluation partner designed and developed an evaluation protocol with the help of the evaluation technical support team. In Case Seven, the stakeholders assisted in determining the timing of when to approach the evaluation participants.

A table was generated (See Table 6.12) to depict the types of human factors that influenced the evaluators' decision making. The first column in the table lists the type of human factors found in the expert evaluators' narratives. The second to the eighth column indicate the numbers of the cases in which the factor types could be found in the narratives. The rows indicated a specific type of human factor and their matching case numbers.

Table 6.12

Human Factors

Human Factor	Case Number						
	1	2	3	4	5	6	7
Evaluation Team Members' Attitudes Toward One Another				x	x	x	x
Evaluator's Attitude Toward Stakeholders	x	x	x				
Evaluator's Attitude Toward the Evaluand	x						x
Evaluator's Role	x	x	x	x	x	x	x
Stakeholders' Attitudes Toward		x	x	x	x	x	

Human Factor	Case Number						
	1	2	3	4	5	6	7
One Another							
Stakeholder's Attitude Toward Being Evaluated	x	x		x	x		x
Stakeholder's Attitude Toward the Evaluator	x	x					
Stakeholder's Attitudes Toward the Evaluand	x	x	x	x			x
Stakeholder's Role	x	x	x	x	x	x	x

Real-World Factors

According to Orasanu and Connolly (1993), there are "key contextual factors that affect the way real-world decision making occurs." The factors were identified as action/feedback loops; high stakes; ill-structured problems; uncertain, dynamic environments; shifting, ill-defined, or competing goals; multiple players; organizational goals and norms; and time stress. The results indicated that all of the factors influenced at least seven of the cases, and changes were made to the list. For example, data analysis indicated that timing was an important real-world decision making factor that influenced the expert evaluators' decision making, and that evaluators not only had to work with multiple participants, but more importantly, their work also involved encounters with multiple stakeholders.

Action/feedback loops. In all of the cases, there were action/feedback loops present in the critical incidents. In seven of the cases, the research participants received feedback for their proposed actions or decisions. In Case Two, the research participant encountered a funding agency sponsoring the evaluation that would not respond to her requests for collaboration in designing the evaluation; therefore, the evaluator relied on her training and experience to guide her actions. However, the evaluator reported that the evaluation report was not well received. The evaluator stated that the funding agency did not consider the evaluation report to be an

"impact evaluation report but rather a site visit." Although, the evaluator tried to explain the findings, she was unsuccessful.

Patton (2008) conducted research on the situation responsiveness to the evaluation environment. The research indicated that "there is no one best way to conduct an evaluation. This insight is critical. The design of a particular evaluation depends on the people involved and on the situation" (p. 199).

High stakes. All of the research participants indicated that the evaluations were high stakes because the evaluations provided recommendations that would improve or inform a program that was addressing a life-and-death issue. For example, the expert evaluator in Case One indicated that the evaluation was important because the program addressed HIV, a life-and-death issue. The program's poor performance and its possible negative impact on persons living with HIV was one of the reasons why he felt empowered to confront the health department official who resisted being evaluated.

Ill-structured problems and uncertain, dynamic environments. Orasanu and Connolly (1993) list the "Ill-structured Problems" and "Uncertain, Dynamic Environments" as two "characteristics that affect the way real-world decision making occurs" (p. 5). These characteristics were found to be intertwined, such that it was not possible to separate them. Hence, they are listed in a single category. All of the cases exhibited aspects of ill-structured problems and uncertain, dynamic environments.

For example, in Case Two, the evaluator was uncertain about the reasons as to why the evaluation fund agency was not engaged in the evaluation design process. The evaluator decided to use an exploratory approach because of the funding agency's lack of engagement. Thus, her design evolved as she discovered how to implement the evaluation. The uncertainty continued

as she tried to understand why the lead consortium did not organize the interviews and focus groups as agreed. The evaluator relied on her own assumptions in determining that the lead consortium was avoiding being evaluated. In all of the cases, the evaluators' critical incidents involved solving ill-structured problems.

Shifting, ill-defined, or competing goals. In the seven cases, there were shifting and/or competing goals. For example, in Case One, the research participant's goal was to conduct an interview. However, according to the evaluator, the interviewee was trying to avoid being evaluated. The research participant proposed several reasons as to why the interviewee was acting in a hostile manner. Further, he relied on his own interpretation of the official's behavior to determine that the official's goal was to avoid being evaluated.

In Case Five, the research participant's narrative revealed that the conflict existing in the evaluation setting was due to the evaluator's and the stakeholder's opposing goals. The research participant's goal was to design and implement a rigorous evaluation. This, according to the evaluator, conflicted with the organization's goal to have the evaluation conducted on a pre-established framework. The evaluator found himself "trapped" in the organization's "pre-cooked" evaluation design.

Multiple stakeholders. Orasanu and Connelly (1993) listed "multiple players" as a real-world characteristic. In the critical incidents of the seven cases, the research participants worked with multiple stakeholders, who were not only players in the evaluation scenarios but also had a stake in the evaluation outcome. As indicated by Spooner, Flaxman, and Murray (2008),

> The conflicting agenda and perspectives that individual stakeholders and groups bring to an evaluation can impede its effectiveness. The more people are involved in a project, and the more deeply they are involved, the more ideas and agenda there are to be managed. Accordingly, communication is even more important in larger research projects with large numbers of stakeholders. (p. 30)

For example, in Case Four, the evaluator reported that there were conflicting agendas and perspectives between the evaluation partners, which led to a communication breakdown. In Case Two, the evaluator stated that there was also a lack of communication between the evaluation participants, the program funding agency, the lead consortium, and the evaluator.

Organizational goals and norms. As stated by Volkov and Baron (2011), "Evaluation's interface with organizational practices and decision making is largely affected by the organizational climate" (p. 103). The goal is "to create a robust evaluation enabling organizational environment" where there is "a willingness, perhaps even a desire, to find out how well programs have done" (Chelimsky, 2001, p. 23). In addition, Sanders (2002) indicated that

> the problem of developing an evaluation culture in an organization remains perplexing … For those of us who are true believers in the benefits of evaluation, the question of what we can do to make it part of the everyday life of organizations remains unanswered. (p. 253)

In all of the cases, the evaluators needed to work within their or the evaluand's organizational goals and norms. For example, in Case One, the evaluator stated that the government had commissioned the evaluation because it was a part of the donor's funding requirements. In addition, the evaluator needed to work with the health department official. The latter's evaluation resistance was partly influenced by the department's dysfunctional bureaucracy and the government's unclear political positions.

Timing. Orasanu and Connelly (1993) did not list Timing as a real-world problem characteristic. However, it was found to be an important characteristic that influenced the evaluators' real-world decision making. The point in the program's life cycle at which the evaluation was conducted was found to influence the evaluators' approaches and decisions. Case Six provided the clearest example of how the timing of an evaluation could influence the

evaluator's approach in conducting the evaluation. In Case Six, the evaluator decided to conduct an evaluability assessment for determining the evaluand's readiness for evaluation. The evaluability assessment was conducted because the program was to be evaluated early in its life cycle. However, the program did not have established indicators nor did the stakeholders agree on the program's logic model. Moreover, there was a lack of understanding about the program's theory of change, as well as confusion about how the program was being implemented.

Time. In six of the seven cases, it was observed that time stress was present. For example, in Case Five, the research participant stated that the evaluation was commissioned close to the end of the project life cycle. Bamberger et al. (2006) conducted research on how to overcome time constraints when conducting an evaluation. Case Seven was the only case where the evaluator indicated that time stress was not present. Instead, in Case Seven, the evaluator stated that sufficient time assisted him in implementing the evaluation.

Table 6.13 was developed to depict the real-world factors that influenced the evaluators' naturalistic decision making. The table also listed the case numbers in which the real-world factors occurred. The first column includes the type of real-world factor, while the second to the eighth column list the cases in which the real-world factors occurred. The first row after the table headers lists a specific type of real-world factor and in which case(s) the factor occurred.

Table 6.13

Real-World Factors

Real-World Factors	Case Number						
	1	2	3	4	5	6	7
Action/Feedback Loops	x	x	x	x	x	x	x
High Stakes	x	x	x	x	x	x	x
Ill-Structured Problems and Uncertain, Dynamic Environments	x	x	x	x	x	x	x
Shifting, Ill-defined, or Competing Goals	x	x	x	x	x	x	

Real-World Factors	Case Number						
	1	2	3	4	5	6	7
Multiple Stakeholders	x	x	x	x	x	x	x
Organizational Goals and Norms	x	x	x	x	x	x	x
Timing	x	x	x	x	x	x	x
Time Stress	x	x	x	x	x	x	

Research Question Three:

How do expert evaluators make naturalistic decisions when evaluating HIV/AIDS health education programs?

The results indicated that the expert evaluators' naturalistic decision making process required the use of situation assessments, philosophical assumptions, interpersonal skills, and satisficing. Two decision making process cycles were identified based on the results of the analysis of the decision making processes used by the evaluators.

Situation Assessment

The research participants relied on situation assessments (Crandall, Klein, & Hoffman, 2006; Klein, 1993) to assist in conducting the evaluations. An initial assessment assisted in understanding the evaluation context and the encounter with the critical incident(s). Subsequent situation assessments were utilized throughout the decision making process.

The situation assessments involved comparing the critical incident(s) to previous similar incident(s) [prototypes]. The research participants relied on two types of prototypes: personally experienced prototypes and evaluation peer based prototypes. The peer-generated prototypes were generally shared through stories, literature reviews, or published articles from the evaluation community. The prototypes were utilized to assist in the identification of expectations, relevant clues, plausible goals, and typical actions (Crandall, Klein, & Hoffman, 2006; Klein, 1993).

The responses to a critical situation, also referred to as "situational responsiveness" by Patton (2008, p.28), were informed by the existing theories, logic, and reflection (Kundin, 2008, 2010) to help in identifying the cues, assigning meaning to cues, and utilizing the meaning found to predict their relevance to the current or future events (Crandall, Klein, & Hoffman, 2006; Klein 1993). The theories (Dillon, Buchanan, Corner, 2005; Greene, 2005), logic, reflections (Crandall, Klein & Hoffman, 2006; Klein, 1993; Simon, 1955, 1956, 1973), and judgments were the informed personal experience or knowledge that existed in the evaluation community (Kundin, 2001). If a prototype could not be discerned, then a situation assessment was used to guide research participants' situational responsiveness.

Comparison to personally experienced prototype. It was observed that the research participants' situation assessments followed a pattern. First, they determined whether they had had a personal experience that would inform them about what was occurring. Then, they compared the current situation to what they expected to happen, the cues that indicated if the expected occurred or not, the likelihood that their goals would be achieved, and the typical action taken in a similar situation. The knowledge and experience utilized for understanding the situation was based on personal knowledge or experience. The evaluators also relied on the knowledge gained from interactions with the evaluation community or their peers. The evaluators provided a description of the mental simulations or imaginations about what they anticipated would occur. Then, the evaluators shared their reflections, logic, and theories as to what and why something was occurring.

Case One provides an example of the evaluator's reliance on previous experience [prototype] to identify his anticipated expectations when conducting an interview with the officials. He also shared his reliance on a past evaluation experience with an expert evaluator to

inform him on what decisions were the most appropriate and feasible. The evaluator also relied on prototypes to help him determine what resources were available [such as his authority] and anticipate the possible outcomes of his decisions [further resistance to being evaluated or cooperation]. He relied on his judgment to determine which decision to choose as well as when to implement it. For example, the evaluator first attempted to explain the purpose of the interview. Next, he tried to reframe her concerns. Finally, the evaluator made an effort to use his authority as the voice of the organization that was funding the intervention. He relied on his judgment to determine which prototype best matched the situation. The evaluator was confident about his decisions because of his previous experience(s), knowledge, and judgment(s). He utilized reflection to assist him in making decisions and determining the possible outcomes of this decision. For example, he reflected on "not going over her head and go to her boss" because it could become abusive, which could further infuriate the hostile official. He also wanted to avoid abusing his power.

Case Six depicts another example when the evaluator was trying to figure out why the evaluation team was being uncommunicative. She relied on her previous experience knowledge and utilized questions to understand the situation. She stated,

> One can have many experiences from different countries. For example, when you are in-country, everybody is excited about the project but as soon as you come back to headquarters overseas, the project is at a standstill. There are so many reasons as to why that's so. It is good to ask yourself, "Why is that the case?" Instead of making an assumption, you try to know *why* … For example, when I send an email and there isn't a response, I wonder "Maybe you didn't read it? or "Maybe you didn't agree."

Comparison to peer based prototypes. In Case Four, the expert evaluator utilized structured interviews to obtain the information necessary to secure entry into the community to be evaluated. The evaluator stated that she expected structured interviews to assist with

understanding the evaluand context, not only based on her experience but also on what she has learned through her training and the social science literature.

Decision Making Cycle: Pattern One and Pattern Two. Decision making depends on the context in which the evaluation decisions are made as well as the types of decisions to be made (Owen, 2007). The evaluators' decision making process followed two patterns. Pattern One followed the following format: an evaluator encountered a critical incident (a problem), he or she conducted an initial situation assessment, generated a decision, received stakeholder feedback, conducted ongoing situation assessments as he or she evaluated the stakeholder feedback, and determined if the problem was solved. The number of cyclic iterations depended on how quickly the critical incident was resolved. The number of iterations in Pattern One was dependent on how many cycles needed to occur to arrive at a decision (see Figure 5.2).

Unlike Pattern One, in Pattern Two the evaluator did not receive feedback from the evaluation stakeholder. Pattern Two began when the evaluator encountered an incident, generated a decision, tested the decision, did not receive stakeholder feedback, and relied on his or her experience and training to reach a decision. The evaluator decision making cycle involved a single iteration for each decision that was generated due to a lack of feedback.

For example, in Case Two, the evaluator reported that she did not receive feedback when trying to engage the evaluation sponsor so that the evaluation approach, questions, and evaluation participants could be determined. The evaluator stated,

> I tried to call [the donor] before I went out to the field because people define 'impact' differently … [and] I needed to know 'What they were trying to find out?' I can focus evaluation questions if they tell me what they want, why they are doing it, and who is going to use it. [That way] I can say [to myself], "Look, I know the donor really wants to know X, so I need to interview those people."

Thus, the evaluator relied on her experience and training. In the narrative, she indicated that she "needed to make assumptions and decisions in the field about what data to go after … [and] make assumptions [experience and training] about the focus of the evaluation," as the donor would not return her calls. The evaluator said that she used an "exploratory approach," as vital information was found to be missing.

Philosophical Assumptions

The data analysis indicated that evaluators relied on their philosophical assumptions to inform their situation assessments. Mertens (2008) found that evaluators struggle with four philosophical assumptions when conducting an evaluation: axiological, ontological, epistemological, and methodological assumptions. Kundin's (2008, 2010) also found that evaluators relied on theories, logic, and reflection to assist them in conducting a situation assessment. Zsambok and Klein (1997) conducted research on how people made decisions under conditions of "time pressure, ambiguous information, ill-defined goals, and changing conditions" (p. 287), which are elements in naturalistic decision making. Their research indicated that decision makers rely on expectations, plausible goals, relevant clues, and typical actions.

The tables depicting the evaluators' use of theories-in-use in the real world, logic and reflection goals, results, and relevant cues in the decision making process are very lengthy. Case Four was selected as a representative table, as it encountered multiple critical incidents with multiple decisions. Understanding the Evaluand's Context depicted a situation assessment with multiple decisions. The elements in the table were depicted, as much as possible, in the same order as they appeared in the narrative.

In Table 6.14, the first column lists the elements used in the situation assessment of a critical incident that the evaluator encountered. The second column reports the quotes that match

up with the elements. There were four table spanners embedded into the table. The table spanners were as follows: Evaluation Context, Decision 1, Decision 2, and Decision 3. Descriptions of the decisions made by the evaluators are provided in the table. The rows contained specific information about the elements listed under the table spanners.

Table 6.14

Case Four Situation: Assessment Understanding the Evaluand's Context

Element	Quotes
	Evaluation Context
General Logic: Understand Evaluation Context and Two Countries' Populations to Secure Successful Community Entry	The evaluator in this scenario was trying to secure community entry in two countries so that the team could conduct an ex-ante evaluation: "Before we could officially submit our proposal, we needed to make sure that we followed our own organization's internal requirements, such as writing a protocol, which was a major endeavor."
Plausible Goals: Gather Data	The evaluator stated that to "conduct a successful evaluation you need to gain an understanding of the culture, context, and the changes in the context during the evaluation …"
Theories-in-Use in the Real World: Judgment about the Importance of Planning Community Entry	The evaluator stated that "in Sub-Saharan Africa, I observed other studies that didn't recognize the importance of planning entry into the community."
Relevant Clues: Community Entry	"I have seen the challenges that have occurred when evaluators did not take time to understand the context — what it takes to get a study underway … We want to stay on top of any potential problems, rumors, or miscommunications."
Theories-in-Use in the Real World: Protocol for Two Countries is a Major Endeavor	"…we needed to make sure that we followed our own organization's internal requirements, such as writing a protocol, which was a major endeavor."
Working Logic: Different Processes for Different Countries	The evaluator said that "this was because multi-country protocols can be quite laborious because you need to work with both countries." The evaluator stated, "For this evaluation we needed to keep in mind that the different countries had different processes, so for each country we needed to go through different steps."

Element	Quotes
Theories-in-Use in the Real World: Importance of Understanding Evaluand's Context Based on Field Experience and Literature	The evaluator felt that it was "a basic fundamental evaluation principle [to understand the evaluand's context] … [that] has been verified by my [the evaluator's] field experience … as well as what I have read about evaluation and the social behavioral sciences."
Decision 1: Conduct a Literature Review and Field Visits	
Theories-in-Use in the Real World: Conducting a Literature Review	The evaluator stated that for this evaluation, the team conducted "a thorough literature review" for each of the countries, and the team decided to visit "the two countries at least three times … We talked to the ministries in each of the countries and all of the implementing partners … We also asked them to help us identify the populations or country context they were most interested in working with and exploring."
Theories-in-Use in the Real World: Working with Health Departments and Ministry is Challenging	"There have been occasions where getting the health departments on board was challenging … the challenging part was that we needed to sit with them and explain how our proposal relates to their broader mission. It is an important part of the process that you need to plan for."
Working Logic: Relate Evaluation Proposal to Department's and Ministry's Mission	"We are required to make sure that we are aligned with their [ministry's and department's] mission … Sometimes it goes very quickly, sometimes it is slower than you would like. Sometimes there are competing priorities in the ministry, or the key people that you needed to talk to are not available."
Decision 2: Conduct Structured Observations	
Theories-in-Use in the Real World and Relevant Clue: Literature Review is Not Sufficient to Understand Evaluand's Context	The evaluator stated that during the development of the protocol "it became clear, through the literature review and through subsequent discussions that we needed a lot more information on the women's economic, social, and sexual networks."
Theories-in-Use in the Real World: Structured Observations Are a Valuable Data Gathering	The evaluator stated that "[the evaluation team's] decision to conduct structured observations is supported by the literature" and by knowing "what observations can bring to the table."

Element	Quotes
Tool	
Plausible Goal: Understand What Is Occurring in the Population Being Studied.	The goal was to understand what was occurring in the population being studied. "We wanted to understand what was occurring where they work, live, and travel."
Working Logic: Conducting Interviews and Identifying Stakeholders to Understand Evaluand's Context	"… we interviewed or identified people that could inform us of the things we need to be aware of …"
Reflection: Determine Which Questions to Ask Evaluation Participants	"So, when we talked to people and ask them about a study or research we asked ourselves 'Do they know what those words mean?' During the observations we tried to look at what was available, such as 'Is the internet available in the community? What are the agricultural cycles?'"
Decision 3: Postpone Evaluations	
Theories-in-Use in the Real World: Unsafe Conditions	"Due to civil unrest it is not safe to conduct evaluation."
Theories-in-Use in the Real World: Ministry Resisting Human Subject Research	"… just about the time when we were ready to start the evaluation … the ministry made a ruling that there was no basis to do human subject research in the country at all."
Result: Country One	"… unfortunately the evaluation had to be put on hold due to civil unrest."
Result : Country Two	"… the formative evaluation was postponed"

The evaluators' philosophical assumptions indicated that the evaluators used theory and logic during a situational analysis. They identified and performed comparisons with similar situations or prototypes. Decision makers based their situation assessments on the following elements: expectations, typical actions, relevant clues, and plausible goals. Data analysis resulted in the discovery that the evaluators' expectations and typical actions were based on their theories. As a result, the element labeled Expectations was merged under the category labeled

Theories-in-use in the Real World. Similarly, the element labeled Typical Actions was merged under the category labeled General Logic.

Inter-Personal Skills

The data indicated that the interactions between the evaluators and stakeholders affected the evaluators' decision making. It was observed that interpersonal skills were used in decision making. These skills included the following: Communication, Collaboration, Cooperation, and Negotiation. As this is a new category, a brief background on the influence of interpersonal skills in evaluation is provided. Evaluation literature regarding interpersonal factors focused on evaluation utilization. In 1998, Johnson reviewed the literature and models based on evaluation utilization. In the study, Johnson analyzed 17 models and summarized the findings. Johnson stated

> Participation by program stakeholders is essential and continual (multi-way) dissemination, communication and feedback of information and results to evaluators and users (during and after a program evaluation) helps increase use by increasing evaluation relevance, program modification, and stakeholder ownership of results. Evaluators, managers, and other key stakeholders should collaboratively employ organizational design and development principles to help increase the amount and quality of participation, dissemination, utilization, and organizational learning. (p. 104)

A similar and more recent study was conducted by Johnson, Greenseid, Toal, King, Lawrenz, and Volkov (2009). In their study, they researched and analyzed the empirically-based literature published between 1986 and 2005. The study established "that engagement, interaction, and communication between evaluation clients and evaluators is key to maximizing use of the evaluations in the long run" (p. 389). Patton's research on utilization focused evaluations also indicated that successful communication, negotiation, cooperation, and collaboration are necessary if the evaluator needs to avoid or mitigate the impact of evaluation resistance, increase stakeholder participation, or increase evaluation use (Patton, 2008).

Communication skills. In Case One, the evaluator utilized communication and reframing. Walzlawick, Weakland, and Fisch (1974) define reframing as "to change the conceptual and/or emotional setting or view point in relation to which a situation is experienced and to place it in another frame which fits the 'facts' of the same situation equally well or even better, and thereby changing its entire meaning" (p. 95). In Case One, the evaluator "reframed" the evaluation resistant interviewee's concern to show empathy and as an effort to persuade the interviewee to participate in the evaluation process. Kniker (2011) indicated that evaluators not only need methodological skills but also interpersonal skills. Further, the importance of communication was highlighted in the research conducted by Kniker. He mentioned his evaluation experiences where the evaluators hired by government organizations "are technically proficient, but are so focused on the methodological that they miss the practical or have difficulty … conveying data to decision makers in ways decision makers need" (p.69). Kniker indicated that these type of evaluators "don't have the communication skills to work well" (p. 69).

Collaboration skills. In Case Seven, collaboration was a factor that influenced the evaluator's decision making. Unlike the cases in which factors were a barrier, Case Seven indicated that collaboration among the evaluation team members assisted the evaluation process. Collaboration involved making decisions regarding the design and development of the evaluation instruments, as well as when to interview the evaluation participants.

Cooperation skills. A lack of interest or support for evaluation is an evaluation obstacle (Sonnichsen, 2000). In Case One, the evaluator tried to encourage a department official's cooperation during an evaluation interview. The evaluator said that he tried to secure the interviewee's cooperation because the official was "accountable to the donors and the

government tax payers who funded the program." The interviewee's lack of cooperation influenced the evaluator's decision making throughout the evaluation.

In Case Two, the evaluator tried to secure the donor agency's engagement in the evaluation design and obtain the lead consortium's cooperation in the data gathering process. The evaluator decided to use an exploratory approach when conducting the evaluation, due to the donor's lack of engagement. The lead consortium's evaluation resistance —, which manifested itself in a lack of cooperation — resulted in a scramble to schedule interviews and focus groups.

Negotiation skills. "Negotiation is a process of combining conflicting positions" into a non-conflicting one through a process of exchange (Nikolaev, 2007, p.4). According to Patton (2008), conducting an evaluation involves "negotiating with the evaluation's intended and desired users, and adapting the design to financial, political, timing, and methodological constraints and opportunities" (p. 199). In all of the cases, the evaluators utilized negotiation to assist them in overcoming the critical incidents they encountered.

In Case Three, the evaluator had "hope[ed] to conduct a randomized control study," but the NGO working with the evaluation team was "had already made promises as to who would participate in their program." Due to the promises that were made, "conducting a randomized control study was not possible." At first, the evaluation team asked the NGO if there was any "way we [the evaluation team] could get around it." However, the NGO stated that not following through with their promises would negatively impact their credibility. The evaluator was unable to change the NGO's decision to stand by their "promise." As a result, the evaluator decided to conduct a non-randomized comparison study.

In Case Five, the evaluator was unsuccessful in negotiating with the agency that had commissioned the evaluation. The evaluator stated that he was not able to provide any input into

the evaluation design or negotiate for a larger budget and more time. Even though the evaluator's negotiations failed, he conducted the evaluation. The evaluator recommended a thorough evaluation due to his concern about the bias imbedded in the evaluation design.

Self-Determination

Deci and Ryan's research on motivated behavior led to the formulation of self-determination theory (Deci & Ryan, 1985; Ryan and Deci, 2000). According to self-determination theory, people have a need to feel competence and motivation (Gagné & Deci, 2005). Baumeister and Leary (1995) identified relatedness as another need.

Self-determination theory is a macro theory of motivation that is concerned with motivation, individuals' choices, and the extent to which an individual's actions are self-motivated and self-determined **(**Gagné & Deci, 2005). Evaluation has a history of evaluators with a "can-do attitude" to conduct evaluations "where situations are not conducive to traditional qualitative research, including the quasi-experimental (Campbell & Stanley, 1963) and naturalistic methods (Guba, 1978)" (p. 43). The types of self-determination identified as attributes utilized by the evaluators were Autonomy, Motivation, Internal Control of Reinforcement, Competence, Emotional Self-control, and Relatedness.

Autonomy, internal motivation, and internal control of reinforcement. According to Gagné and Deci, "Autonomy involves acting with a sense of volition and having the experience of choice" (2005, p. 333). Gagné and Deci also found that autonomy and motivation are linked to each other:

> Intrinsic motivation involves people doing an activity because they find it interesting and derive spontaneous satisfaction from the activity itself. Extrinsic motivation, in contrast, requires an instrumentality between the activity and some separable consequences … so the satisfaction comes not from the activity itself but rather from extrinsic consequences to which the activity leads. (2005, p. 331)

According to Rotter (1990),

> internal versus external control of reinforcement refers to the degree to which persons expect that a reinforcement or outcome of the behavior is contingent on their own behavior or personal characteristics versus the degree to which persons expect the that the reinforcement or outcome is a function of chance, luck, or fate, is under the control of powerful others, or is simply unpredictable. (p. 489)

Stevahn, King, Ghere, and Minnema (2005) found that "[p]rogram evaluation began as a field with a can-do attitude." The evaluators attempted to successfully implement the evaluations in spite of the challenges that they encountered. For example, in Case Five, the inability to influence the funding agency could easily make the evaluator believe that he would be unable to conduct the evaluation. Instead, the evaluator found other funding resources and completed the evaluation.

In Cases One and Two, the evaluators encountered evaluation resistance. In Case One, the evaluator did not succumb to the influence of the more "powerful other" — the government official who refused to be interviewed. Instead, the evaluator continued in his efforts to persuade the interviewee to participate in the evaluation interview. In Case Two, the evaluator did not stop the evaluation process even though the funding agency and the lead consortium were barriers to conducting the evaluation. Instead, the evaluator decided to conduct the evaluation using an exploratory approach and persuaded the lead consortium to assist her in securing the interviews and focus groups that were needed to gather evidence for the evaluation.

Competence. McClelland and Boyatzis (1980) defined competencies as "a generic body of knowledge, motives, traits, self images, and social roles and skills that are causally related to superior or effective performance in the job" (p. 369). Cognitive evaluation theory suggests that "feelings of competence as well as feelings of autonomy are important for intrinsic motivations" (Gagné & Deci, 2005, p. 332).

Stevahn, King, Ghere, and Minnema (2005) defined program evaluation competencies as a "set of unique skills and knowledge that distinguishes professional evaluators" (p. 230). Stevahn et al. (2005) indicated that the "development of competencies for evaluators — [in] the field of program evaluation … has yet to happen" (p. 230). King, Stevahn, Ghere, and Minnema (2001) developed a taxonomy of essential evaluator competencies, aimed at generating an agreement that "assist[s] professionals in the field to structur[e] the following: training programs for novice practitioners; continuing education programs for experienced professionals; periodic review to update the competencies as theory, research and practice" (p. 44).

In Case Two, the evaluator indicated that training and field experience are important factors when an evaluator conducts an evaluation. He stated

> One of my research interests is evaluator competencies. I struggle with that a lot. I think it's *extremely* important to understand evaluation theory and program theory … however, if that is not equally matched with field experience, I'm not sure you are going to be a top-notch evaluator.
>
> I've been doing evaluation for almost 19 years. Evaluation isn't just about getting a degree in evaluation or research, though I think it helps to have that kind of training. When I was first doing evaluation, I was not bringing the wealth of experience that I bring now. I think it is important that people know that I've been in the field for that amount of time. I don't sit in the United States and fly out to Africa once in a while to do an evaluation. I've been in the field this whole time. I think it makes a big difference being in the field. It puts you more in touch with the reality of what I am evaluating.

Emotional self-control. In six of the seven cases, the evaluator exhibited emotional self-control. For example, in Case One the evaluator utilized self-control until he finally decided to "push back" against the interviewee's resistance to being evaluated. The expert evaluator said, "I generally come across quite easy, non-confrontational person, but when I do get irritated, there is a clear change in character." The evaluator decided to confront the interviewee using "firm" tones.

In Case Two, the evaluator indicated frustration and anger toward the two primary stakeholders. The evaluator expressed frustration when the agency responded angrily to the evaluation report. The evaluator tried explaining that the evidence indicated that the use of a consortium model was not working. The evaluator expressed self-control when she did not express her thoughts: "I could have given them an impact report, but it would have been only one paragraph long."

Relatedness. Ludema and Virgilio (2007) described relatedness as "the need people have to feel a sense of belonging and connectedness with others. When people have a sense of relatedness, they feel like they are making a contribution to the greater whole and that the greater whole is making a contribution to them" (p. 27). It was observed that all of the expert evaluators expressed relatedness. For example, in Case Two, the evaluator indicated that she primarily considered herself to be a development worker whose aim was to better the lives of people in developing countries. Despite the challenges she encountered, the evaluator expressed a sense of satisfaction when she said

> Even this experience had some useful stuff come out of it. We did manage to run one focus group with a bunch of women that were participants in the intervention. They said, "Nobody listens to us." This was the first time they had someone come out, sit down with them, and listen to them talk. That's great. The fact that I was able to talk to the different consortium members and say, "You know, this is not evaluation. Let me tell you about evaluation" — I think that it is a happy ending. They understood that evaluation can be so many things that could be helpful and useful to them — not just the donor coming out or sending someone out for five days. So, that's a happy thing.

In Case Six, the evaluator expressed relatedness when she communicated excitement in her voice regarding the "empowerment" that the evaluation partner felt after she successfully broke down the communication barriers and provided the leadership training.

Satisficing. Satisficing is defined by Simon (1955) as a decision making strategy in which a decision maker generates a decision that is adequate to solve a problem based on criteria instead of generating an optimal decision. In the seven cases, the expert evaluators used satisficing when they attempted to solve a problem. The decisions were the best the evaluator could do, "given the situation." It was observed that levels of satisfaction varied with their decisions.

An example of satisficing can be found in Case Two. The evaluator decided to use an "exploratory approach," due to the funding agency's lack of engagement. However, the evaluator indicated that it was not an optimal approach for conducting an evaluation. She stated that "university professors might frown on the methodology that I ended up using" [but] "I had to stay [to complete the evaluation]."

Another example of satisficing can be found in Case Five. The expert evaluator stated that the sponsor of the evaluation did not provide him an opportunity to provide any input into the design. The evaluator was not able to negotiate the methods used in the evaluation, the evaluation questions, members of the evaluation team, sites that the evaluator could visit and observe, budget, and length of time needed to complete the evaluation, due to the evaluation sponsor's inflexibility.

The evaluator also stated that he was unable to present the evaluation findings in a preferred manner. However, the evaluator had to provide the evaluation sponsor with a "short report." The evaluation team did not provide a "formal evaluation presentation … because it required another visit, which adds to the time and budget of an evaluation." The evaluator indicated that he would have preferred to share the findings in a "formal evaluation presentation" because he considered it "a very good [evaluation] practice."

Variation of Critical Incidents, Factors, and Decision Making

In my study, I used phenomenography to further understand how evaluators made naturalistic decisions. Phenomenography focuses on the interpretation of the phenomenon — studied from the research subject's point of view—and interprets the qualitatively different ways in which the research subject experiences a phenomenon. I analyzed the cases to discover if there was any variation among how the evaluators reported: the critical incidents they encountered, the factors that influenced their naturalistic decision making, and how they made naturalistic decisions.

Critical Incident

This section reports the results of data analysis that focused on the types of critical incidents that the evaluators encountered. A review of this study's seven cases resulted in the findings that the evaluators reported critical incidents as a combination of barriers or enablers and expected or unexpected events. Table 6.15 depicts the variation among the critical incidents that the evaluators encountered.

Table 6.15

Critical Incidents: Barriers or Enablers, and Expected or Unexpected

Critical Incidents	Case Number			
	Barriers Expected	Barriers Unexpected	Enablers Expected	Enablers Unexpected
Barriers to Communicating with Stakeholders		4, 6	7	
Barriers to Community Entry	4			7
Barriers to Coordinating Stakeholders	4			
Barriers to Understanding the Evaluand's Context	4	6	7	
Budget Constraints		3, 5, 6		
Data Constraints		3, 4, 6		
Evaluation Resistance		1, 2		

Lack of Barriers to Collaborating with Stakeholders		7
Threats to the Evaluation Design	2, 3, 5, 6	
Time Constraints	4, 5,	7

For example, during the critical incident labeled “Barriers to Understanding the Evaluand's Context,” the expert evaluator stated that the Ministry’s decision to halt human subject research was a barrier and unexpected critical incident. The evaluator said, “In the other country where the formative evaluation has been postponed, just about the time when we were ready to start the evaluation, the Ministry made a ruling that there was no basis to do human subject research in the country at all.” She also indicated that it was unexpected when she stated, “I consider that an unexpected barrier to community entry.”

The critical incident in the category labeled “Barriers to Understanding the Evaluand's Context” was reported as a barrier and expected critical incident. The evaluator in Case Four indicated that understanding the evaluand’s context was a barrier when she said, “To understand the context where the evaluations are to be conducted, we conducted structured observations.” The evaluator indicated that the critical incident was expected when she said, “The structured observations were part of the official protocol.”

There were instances where enabling and unexpected critical incidents occurred. For example, in Case Seven, the evaluator stated that the community participants supported the circumcision initiative when they viewed circumcision “as a medical procedure” rather than a “religious ritual” imposed on the community. The evaluator stated, “So the religious aspects associated with the surgical intervention to prevent the transmission of HIV/AIDS seemed to have worked themselves out.” The evaluator also reported the expected challenges when he tried to use data gathering instruments from another country with a different culture than his current

evaluation environment. He said that he understood that data instruments used in one culture could not be automatically used in another culture.

Factors That Influenced Naturalistic Decision- Making

In their narratives, the evaluators indicated factors that negatively or positively influenced their decision making and as a result influenced the evaluation outcomes. For example, in Case Three, an insufficient budget influenced the evaluation team's decision to hire an NGO to gather data for the evaluation. The data was not usable, as the NGO did not have the skills needed to gather high quality data. Thus, some of the evaluation questions could not be answered. This had a negative impact on the evaluation findings and the report.

In contrast, factors can have a positive influence on evaluator's decision making. For example, in Case Six, concerns about evaluation utility guided the evaluation team's decisions regarding which evaluation questions should be addressed. Furthermore, concerns about "evaluation use" influenced the evaluation team's decision to provide evaluation capacity building opportunities.

It was observed that not all negative factors resulted in negative evaluation outcomes. For example, in Case One, politics had a negative influence on the initial interaction between the evaluator and the evaluation participant. The evaluator indicated that the volatile political environment was one of the reasons why the evaluation participant resisted being interviewed. On the other hand, the evaluation participant's acquiescence to the prevailing political environment influenced the evaluator's decision to assertively pursue the interviewee's participation in the interview. Thus, the evaluator's persistence resulted in a complete interview that provided the necessary data for the evaluation. It was observed that all of the positive factors did result in positive evaluation outcomes.

This section reports the findings regarding the variation present in the expert evaluators' decision making when evaluating HIV/AIDS health education programs, and is guided by the research question: How do expert evaluators make naturalistic decisions based on critical incidents they encounter and the factors that influence their decision making when evaluating HIV/AIDS health education programs?

Evaluators' Naturalistic Decision Making

Expert evaluators often make multiple decisions when conducting evaluations. Further analysis indicated that stakeholder feedback influenced the number of decisions the evaluators made. For example, in Case One, the evaluator made an initial decision, requested for stakeholder feedback, received the feedback, and adjusted his decisions until he was able to convince a government official to participate in an evaluation interview. It thus involved three decisions to resolve the situation. Unlike Case Three, the evaluator in Case Two made multiple requests for stakeholder feedback on how to design the evaluation. In spite of her attempts, she did not receive any feedback from the agency funding the evaluation. Thus, the evaluator stated that the funding agency refused to engage in the evaluation design process. The evaluator indicated that she needed to make a single decision informed by her evaluation experience, knowledge, and skills, because she did not receive stakeholder feedback.

Chapter Summary

In this chapter, I described the results of my study. I analyzed the research participants' narratives to obtain the results. The analysis focused on understanding the critical incidents that the evaluators encountered, the factors that influenced the evaluators' decision making, and how the expert evaluators made naturalistic decisions. The results were organized in accordance with the research question's themes, categories, and sub-categories—later referred to as types.

This chapter reported the critical incidents that the evaluators encountered. The critical incidents the evaluators encountered were identified as follows: Barriers to Communicating with Stakeholders, Barriers and Enablers to Community Entry, Barriers to Coordination, Barriers to Understanding the Evaluand's Context, Budget Constraints, Evaluation Resistance, Successful Collaboration with Stakeholders, Threats to High Quality Data, Threats to the Evaluation Design, and Time Constraints. In addition, data analysis identified the factors that influenced the expert evaluators' naturalistic decision making as follows: Context Factors, Evaluation Factors, Human Factors, and Real-World Factors. Further, how expert evaluators made naturalistic decisions was also reported.

My data analysis identified that evaluators conducted situation assessments, relied on philosophical assumptions, used interpersonal skills, relied on self-determination, and utilized satisficing when making naturalistic decisions. Further analysis indicated that expert evaluators reported critical incidents as barriers or enablers and expected or unexpected. Evaluators also reported naturalistic decision making factors as either positive or negative factors. An analysis of how evaluators made naturalistic decisions indicated that the number of decisions evaluators made depended on whether the evaluators received stakeholder feedback or not.

CHAPTER 7

FINDINGS, IMPLICATIONS, AND CONCLUSIONS

This chapter begins by summarizing how I used the research design of my study to conduct my research. The research design utilized CDM, narrative analysis, case studies, and phenomenography to elicit information, represent evaluators' knowledge, and analyze data. The chapter also summarizes the key findings of my study and their implications for research and practice. Lastly, I have also provided the conclusions derived from my study.

Research Design

I began this research with the objective of understanding how expert evaluators made naturalistic decisions when evaluating HIV/AIDS health education programs. To achieve this goal, I conducted semi-structured retrospective interviews to gather stories of expert evaluators' field experiences. I interviewed evaluators with at least five years of evaluation experience. During the interview, I asked them to share stories based on an evaluation experience in which they had encountered critical incidents.

I conducted two one-hour interviews with each participant by utilizing the critical decision method. A literature review revealed that the evaluation context influences how evaluators conduct their evaluations. I decided to generate co-constructed member-checked narratives and decision making timelines to capture the thick and rich descriptions of the evaluation contexts in which evaluators work. I developed these narratives into cases, such that I could study the unique features present in each narrative. Furthermore, I chose phenomenography to help me understand the qualitatively different ways in which the evaluators

described the following: the critical incidents they encountered, the factors that influenced their decision making, and the manner in which they made naturalistic decisions.

Summary of Findings

The framework guiding my research (see Figure 2.3 and 2.4) is based on a literature review of how evaluators plan evaluations, Klein's RPD model (1997), and the framework within which evaluators make everyday practice decisions (Kundin, 2008, 2010). Based on a review of the cases, I determined that evaluators conducted evaluations during the planning, implementation, and end stages of various programs' lifecycles. I also found that evaluators encountered critical incidents at different stages of the various evaluations' life cycles.

During the analysis, I discovered that evaluators encountered the following critical incidents: Barriers to Communicating with Stakeholders, Barriers and Enablers to Community Entry, Barriers to Coordination, Barriers to Understanding the Evaluand's Context, Budget Constraints, Evaluation Resistance, Successful Collaboration with Stakeholders, Threats to High Quality Data, Threats to the Evaluation Design, and Time Constraints. Further analysis determined that the factors influencing evaluators' naturalistic decision making were: Context Factors, Evaluation Factors, Human Factors, and Real-World Decision Making Factors. I found that evaluators conducted situation assessments and relied on their philosophical assumptions, interpersonal skills, and self-determination, in addition to satisficing, when evaluating HIV/AIDS health education programs.

Critical Incidents

When attempting to conduct systematic and objective evaluations, the evaluators used criteria to judge the merit or worth of an HIV/AIDS health education program. I discovered that the evaluators encountered critical incidents that either thwarted or promoted their attempts to

conduct systematic and objective evaluations that judged the merit or worth of a program. I also found that critical incidents occurred at different points in the phases of evaluation life cycles.

Evaluators also experienced events, such as time or budget constraints that threatened the overall evaluation process. In addition, the evaluation teams encountered resistance as well as critical incidents that threatened their ability to enter an evaluand's community and gather high quality data. Moreover, evaluators encountered critical incidents that were barriers to their attempts to communicate or coordinate with the stakeholders. The evaluators also encountered events, which made it difficult to understand the context in which they were working. However, not all critical incidents were viewed as barriers. For example, evaluators reported that having the stakeholders' support, sufficient time to complete an evaluation, and successful collaboration with stakeholders assisted in implementing the evaluations.

The evaluators attempted to be proactive so that they could prevent the occurrence of any negative critical incidents. However, the evaluators could not anticipate all the challenges they encountered. At the same time, they also encountered unexpected events that assisted them with implementing an evaluation, such as unanticipated community and political support for conducting the evaluation.

Technology. I found that the evaluators relied on technology, while conducting the evaluations. The evaluators considered telecommunication as an important medium that bridged the communication gap between the geographically distant evaluation team members. Evaluators used such technology to gather evaluation data, act as a knowledge management system, and track trends. To that end, evaluators conducted resource assessments to determine the technology they needed when they went into the field to conduct evaluations.

Participatory evaluation. While carrying out cross-cultural evaluations, I found that the evaluators conducted participatory evaluations because it increased their ability to establish partnerships. Evaluators also conducted participatory evaluations because it provided greater opportunities to collaborate with the stakeholders (Bamberger et al., 2006; Mertens, 2009). In effect, the evaluators used evaluation approaches to communicate with the stakeholders.

Communication. In one case, the evaluation team used participatory evaluation approaches because the cultural backgrounds and languages of the evaluation teams and the participants differed. To overcome language and cross-cultural barriers, the evaluation teams consulted with the local technical support teams or key stakeholders so that the evaluators could communicate with the evaluation participants and design appropriate evaluation instruments. In addition, the evaluators also relied on the local stakeholders to assist in securing the evaluation team's community entry.

Collaboration. The successful implementation of participation evaluations depends on the stakeholders' active participation in the evaluation process. However, the stakeholders were not always willing to engage in the evaluation process. For example, in one case an evaluator tried to engage the evaluation funder so that the evaluator and the funding agency could collaboratively develop an appropriate evaluation design. The evaluator designed the evaluation without the funding agency's input as the funding agency would not engage in the evaluation design process. However, this turned out to be a problem. When the evaluator reported the evaluation findings and submitted the evaluation report, the funding agency stated that the evaluation did not meet their needs.

Participation evaluation approaches are not limited to involving stakeholders in the evaluation process. Stakeholders also need to allow evaluators to contribute to the evaluation.

For example, an evaluator encountered an instance in which an evaluation-funding agency did not allow an evaluator to provide any input into the evaluation design. As a result, the evaluator decided to recommend that the evaluation funder commission a new evaluation.

Conflict management and evaluation resistance. At times, evaluators encountered evaluation resistance when conducting evaluations. During data analysis, I discovered that evaluators reported that evaluation resistance could occur when evaluators design an evaluation, attempt to secure community entry, gather data, or report their evaluation findings. The evaluators indicated that evaluation resistance occurred because the stakeholders felt threatened by information that might be uncovered during the data collection process (Stevahn & King, 2005) or by the evaluation reports. For example, an evaluator reported that a department official resisted being interviewed because she was concerned that the evaluation findings would have a negative impact on her career.

Capacity building. The evaluators attempted to provide evaluations that were useful and utilized. In one case, an evaluator successfully provided capacity-building activities through the evaluation to increase evaluation participants' understanding of how evaluations can improve programs. The evaluators attempted to design evaluations so that the evaluation-funding agencies would use the information to improve their program. Despite one evaluator's efforts to encourage an evaluation-funding agency's participation in the evaluation design process, the funding agency did not respond to her repeated requests for feedback on the evaluation design. As a result, when she submitted the evaluation findings and report, the evaluation-funding agency was dissatisfied with the evaluation design, report, and findings.

Coordinating multiple stakeholders. An evaluator stated that because the HIV/AIDS health education programs implemented in SSA involved multiple players, coordinating with

stakeholders was an ongoing challenge. As programs and evaluations are becoming more complex (Patton, 2010), coordinating with stakeholders also becomes more complex. Communication and planning assisted the evaluators in their attempts to coordinate with multiple stakeholders.

Evaluation techniques. I found that evaluators adapted evaluation techniques to the specific situations they encountered. Literature reviews, structured observations, and interviews were some of the techniques that the evaluators used to build consensus and understand an evaluand's context before planning a community entry. In one case, an evaluator used a logic model (Wholey, Hatry, & Newcomer, 2004) to build consensus among evaluation partners. In another case, the evaluation team conducted structured observations and interviews to time their entry into a community.

Inadequate Funds. I discovered that evaluators encountered budget constraints that hindered their ability to complete an evaluation. For example, an evaluation was placed on hold due to a non-payment. Insufficient funds also influenced an evaluator's decision to hire a technical support team with the essential research skills and to use a small sample size. In another case, insufficiently funded evaluations led to an evaluator's attempts to negotiate a budget increase. However, the evaluator's efforts to increase the budget were unsuccessful. As a result, the evaluator authorized the use of his department's funds to supplement the budget shortfall.

Proactive efforts. I found that evaluators were taking proactive steps so that they could successfully implement evaluations. The evaluators' inability to predict potential barriers to conducting an evaluation was due to the challenges they encountered when conducting

evaluations in the field. This is consistent with Schwandt and Dahler-Larson's (2005) research, which indicated that evaluators encounter challenges when they conduct evaluations in the field.

For example, when evaluators attempted to secure community entry, they tried to obtain the stakeholders' trust, willingness to cooperate, and collaboration. In spite of the evaluators' proactive efforts, they struggled to secure community entry in two countries. The narratives also highlighted that evaluators were not able to anticipate all the situations they encountered. For instance, in one country, unexpected civil unrest influenced an evaluation team's decision to postpone an evaluation and leave the evaluation site. In another case, a country's ministry decided to restrict all human subject research. Thus, the ministry's decision resulted in the evaluation team's decision to leave the country and postpone the evaluation.

Community and political support. Evaluators encountered events that facilitated their ability to implement evaluations. For example, an evaluator reported that community and political support for an intervention assisted in securing the stakeholders' support for the evaluation. Based on the evaluator's narrative, I discovered that if a program's stakeholders supported an intervention, it was likely that the evaluation would also receive stakeholder support.

Insufficient time. The six evaluators reported that they encountered time constraints because they did not have enough time to conduct an evaluation (Bamberger et al., 2006). They also reported that events occurred that added to their evaluation's timeline. In one case, the evaluator reported that at least six months had been added to an evaluation's timeline because a contractor did not follow a contract's requirements. Thus, the evaluation team had to coordinate a new round of requests for proposals. Similarly, as an evaluator reported, that a year would be added to the time between the submission of an evaluation proposal and its approval when

evaluators worked in organizations with onerous bureaucratic environments as well as when the evaluation sponsors relied on over-worked employees to review an evaluation proposal.

I found that evaluators reported that time constraints were also caused by other reasons, such as rapid data gathering requirements and when the evaluation was scheduled in a project's lifecycle. The evaluators compensated for data gathering time constraints by developing rapid data gathering techniques. One evaluator reported that time constraints occurred because an evaluation had been commissioned at the end of a program's life cycle.

However, I found that not all evaluators encountered time constraints. An evaluator indicated that implementing an evaluation was less stressful because he had sufficient time to complete the evaluation. He stated that having sufficient time helped him coordinate the logistical challenges that often occur when conducting an evaluation.

Factors that Influenced Naturalistic Decision Making

Evaluators made decisions and took actions to mitigate or solve the problems they encountered. I found that the following factors influenced expert evaluators' naturalistic decision making: Context Factors, Evaluation Factors, Human Factors, and Real-World Factors. I uncovered these factors by studying the critical incidents that the evaluators encountered.

During a review of the evaluator's narratives, evaluators were concerned about understanding the context in which they would be working. Kundin's (2008, 2010) research developed five categories to describe the environment in which evaluators work: descriptors, physical features, climate, interpersonal dimensions, and politics. I determined that Alkin's (1985) research on factors influencing evaluation use was similar to my research. Therefore, I found that I could modify Alkin's categories for my study in this manner: Context Factors, Evaluation Factors, and Human Factors. Furthermore, as I was interested in how evaluators

conducted evaluations in real-world settings, I developed a new category, "Real-World Factors." However, I have not found other studies that focus on the research findings of Orasanu et al. (1993) to interpret the real-world factors influencing evaluators' decision making.

Context Factors. Similar to Patton's (2010) theory that evaluators use situation assessment when conducting evaluations, my study found that evaluators used situation assessments so that they could understand the contexts in which they worked. Kundin's (2008, 2010) research identified that culture influences evaluators' decision making. My study further identified the context factors that influenced evaluators' naturalistic decision making as follows: Civil Unrest, A Country's Infrastructure, Weather, and Agricultural Patterns, Culture, Evaluation Participants' Work, Residence, and Migratory Patterns, Geographical Challenges, Language, Politics, and Power. The context factors were grouped into the following themes: People and the Environment, Politics and Power, and Culture and Language.

People and the environment. The narratives provide, for example, an evaluation team's decision to study work, migration, agricultural, and weather patterns to understand the setting prior to implementing an evaluation. The evaluation team decided to study these patterns to help them determine when to schedule the team's community entry so that their presence was not intrusive. The evaluation team also chose to conduct resource assessments to understand the countries' infrastructures as well as determine the available resources.

Politics and Power. Bamberger et al. (2006) posits that evaluators may need to adapt evaluations when they encounter political constraints. I found that politics influenced evaluators' decision making. For example, one evaluator reported that due to social unrest, the evaluation team decided to leave the country. In another instance, an internal evaluator determined that an evaluation was a political exercise geared for protecting the evaluand. The

evaluator managed to complete the evaluation; however, because bias was built into the evaluation design, he decided to recommend an external evaluation. The evaluator also experienced an event when the evaluation's funding agency used their influence to prevent an evaluator from contributing to the evaluation design.

Culture and language. According to Ebbutt (1998), culture and language are intertwined. In my study, the majority of the evaluators conducted cross-cultural evaluations. Evaluators also conducted evaluations in which evaluators and stakeholders did not speak the same language. In some instances, evaluators established partnerships with people who understood the evaluation participants' culture and language. I found that evaluators established partnerships with local stakeholders to ensure that the evaluation would be culturally sensitive and the language used in the evaluation instruments was appropriate when designing cross-cultural evaluations. In the majority of the cases, the evaluators used a participatory evaluation approach to conduct culturally sensitive evaluations.

Evaluation Factors. Alkin's (1983) research indicated that the procedures, processes, and criteria used for evaluating programs could have an influence on evaluation use. I found that the procedures, processes, and criteria used for evaluating programs influenced evaluators' decision making as well. I also found that in addition to concerns about bias and budget constraints, evaluators' concerns about consent, confidentiality, and privacy requirements influenced evaluators' decision making.

Bias. I observed that because evaluators were concerned about bias, they were interested in ensuring that the evaluations were as "objective as possible" (OECD, 2008, p. 5). For instance, in one case, an internal evaluator was concerned about the bias built into an evaluation design, but was obligated to complete the evaluation. His concerns about bias led to his decision

to recommend that the evaluation-funding agency commission a new evaluation. He was successful in convincing the funding agency to hire a "high powered external evaluator" to take a "new look" at the program's value or merit. The evaluator indicated that he gave more weight to the expert external evaluators' evaluations because he considered expert evaluators more objective.

Budget Constraints. Bamberger et al. (2006) developed a "RealWorld Evaluation Approach" (p. 21) to help evaluators conduct evaluations in instances when "budgets are limited" (p. 22). My study provides examples that describe the budget constraints the evaluators encountered. For instance, an internal evaluator's unsuccessful attempts to negotiate a reasonable budget for an evaluation occurred because his employer denied him an opportunity to provide any input into the evaluation design. As a result, the evaluator secured permission to use his department's funds so that he could supplement the insufficient evaluation budget**.**

Consent, privacy, and confidentiality. I discovered that evaluators have different views regarding evaluation participants' consent to participate in a study. For instance, in one case, an evaluator reported that the evaluation's requirements for participant consent were not as stringent as the typical research requirements. The evaluator believed that the evaluation participant was obligated to participate in an evaluation because the participant received funding from the funding agency, thereby being accountable to the funding agency. In another case, an evaluator was very conscious of the human subject research requirements because of her organization's institutional review board policies regarding securing voluntary participation**.**

I found that evaluators recognized that there was a stigma associated with HIV, and therefore they tried to protect the research participants' privacy and personal information. For example, they made efforts to conduct interviews in locations or settings that protected the

participants' privacy. The evaluators also indicated that they took measures to protect the research data so that participants' identities and personal information would remain confidential.

Evaluation utility. I discovered that evaluators made efforts to ensure that the evaluations were useful and utilized. In one evaluator's narrative, the evaluator reported that she took steps to make sure that the evaluation was useful to the evaluation participants. Unfortunately, the evaluation-funding agency decided that the evaluation was not useful.

The evaluator also stated that Patton's theories about Utilization Focused Evaluations influenced how she conducted evaluations. She stated that when she teaches courses on how to conduct evaluations, she emphasizes the importance of evaluation use and utility. She also stated that she has found it challenging to implement utilization focused concepts in the field.

Evaluation criteria. The criteria evaluators used to determine the merit or value of a program influenced the evaluation questions that the evaluators addressed, the data they gathered, and the evaluation findings. For example, in one case an evaluator tried to determine the effectiveness of a program. She discovered that national HIV/AIDS health supervision standards did not exist. This situation made it difficult for her to determine a program's effectiveness as well as which program indicators to use. The evaluator made a decision to formulate the evaluation questions based on her need to determine the effectiveness of the program and gather evaluation data, which would help develop the national standards to determine the effectiveness of current and future programs.

Data quality. The evaluation team, due to insufficient evaluation funding, hired a technical support team that — unknown to the evaluation team — did not have research expertise. As a result, the resultant quantitative data was of low quality. The evaluator decided that he could not use the data and therefore did not include it in the evaluation report. Another

evaluator indicated that she was concerned that a poorly designed protocol with an unrealistic timeline would result in poor quality data.

Life cycle. The evaluators decided on the type of evaluation they were to conduct based on the point in a program's life cycle at which the evaluation was to be implemented (Kusek, & Rist, 2004). The evaluators conducted formative or process evaluations to assist the program stakeholders in managing their programs. The evaluators also conducted summative evaluations toward the end of the program's life cycle to determine the program's impact.

Poor planning. The evaluators explained that donors and program managers' poor planning resulted in the evaluation participants' unwillingness to participate in the evaluations. An evaluator stated that evaluation participants were subjected to multiple evaluations because several agencies with many reporting structures and evaluation requirements funded a single HIV/AIDS health education program in SSA. Evaluators indicated that this was one of the reasons that they encountered evaluation resistance.

Human Factors. Kundin (2008) identified a category, "People." Alkin (1985) referred to this category as "Human Factors" (p. 29). When Alkin conducted his study, he divided the human factors category into two classes: Evaluator Characteristics and User Characteristics. Alkin separated the human factors category in this manner because he was studying the characteristics that influenced evaluation use. My study was interested in the factors that influenced evaluators' decision making when conducting an evaluation. Thus, I found that it was appropriate to divide the Human Factor characteristics into Evaluator and Stakeholders Characteristics. I identified the following evaluator and stakeholder characteristics: attitudes, philosophical assumptions, and roles.

Attitudes. The evaluators and stakeholders' attitudes toward each other and toward the evaluation (especially during the critical incident labeled "Evaluation Resistance") had an influence on the evaluators' decision making. Ideally, evaluators try to implement an evaluation that is objective (OECD, 2000). I determined that evaluators find it difficult to be completely objective as they deal with emotions that emerge during the evaluation process.

Philosophical assumptions. Evaluators' philosophical beliefs also influenced their decision making. In addition, other researchers have also proposed that different allegiances to evaluation approaches (Mertens, 2008) and opinions about the purpose of evaluations (Patton, 2010) influence how evaluations are conducted. Evaluators' reliance on philosophical assumptions may be the reason why the OECD's definition of evaluation includes the words, "as objective as possible" (p. 5).

Roles. The evaluators' and the stakeholder's roles influenced the evaluators' decision making. For example, the evaluator's perception of the stakeholder's role influenced an evaluator's decision to insist on a stakeholder's participation in an evaluation interview. Similarly, the stakeholder's perception of the evaluator's role, for example, getting the stakeholder in political trouble, influenced the stakeholder's hostility toward the evaluator. Thus, the evaluator adapted his decision to address the stakeholder's increasing hostility.

Real-World Factors. Data analysis resulted in a new categorization scheme, modified nomenclature, and added a factor to the real-world characteristics (see Table 7.1) identified by Orasanu et al. (1993). The factors were categorized as "Context," "Goals," "Problems," "Decision Making Characteristics," and "Time." Analysis of the narratives resulted in changing the Orasanu et al.'s terminology from "Multiple Players" —referring to "not making individual

decisions" (p. 5) — to "Multiple Stakeholders." The term "Stakeholders" highlights that the people involved in the decision making process have a stake in the outcomes of decisions.

I found a method in Orasanu and Connelly's (1993) research that emphasized on contrasting real-world factors with the traditional decision making research paradigm. I added the factor labeled "Satisficing" under the category labeled "Decision Making Characteristics." Similar to Simon's (1955, 1956, 1973) and Klein's (1997) research, I found that evaluators used satisficing when making naturalistic decisions.

Orasanu and Connelly identified "Time stress" (1993, p. 5) as insufficient time for tasks. During my analysis of the narratives, I found that a new real-world factor, "Timing" had emerged. The term refers to the moment when an instance occurs or the spacing of events. I categorized the terms "Time Stress" and "Timing" under the category "Time."

Table 7.1

Real-World Factors Categories, Types, and Definitions

Real-World Decision Making Categories	Factors	Compared to a Traditional Decision Making Research Paradigm
Context	Uncertain, Dynamic Environments	Not static; simulated situations
	High Stakes	Not situations devoid of true consequences for the decision maker
Goals	Shifting, Ill-defined, or Competing Goals	Not clear and stable goals
	Organizational Goals and Norms	As opposed to decision making in a vacuum
Problems	Ill-structured Problems	Problems that are not well-structured
	Ill-defined Problems	Problems that are not well-defined
Decision Making Characteristics	Action/Feedback Loops	Not one-shot decisions

	Multiple Stakeholders	As opposed to only involving individual decision making
	Satisficing	As opposed to generating optimal decisions in an attempt to solve a problem
Time	Time Stress	As opposed to ample time to complete tasks
	Timing	As opposed to not dependent on the spacing of events and the time they occur

Note: Adapted from *Naturalistic Decision Making* (p.5), by C. E. Zsambok and G. Klein, 1997, London: Routledge. Copyright 1997 by Taylor and Francis. Adapted with permission.

Evaluators' naturalistic decision making

In this study, I utilized Klein (1997) and Kundin's (2008. 2010) research to understand how evaluators made naturalistic decisions. The evaluators' knowledge based on previous experiences helped them to understand the atypical and complex situations they encountered when conducting evaluations (Klein, 1997; Kundin, 2008, 2010). I observed that the evaluators conducted situation assessments to understand the context in which they conducted evaluations, similar to Klein's (1997) RPD model and Patton's (2008, 2010) theories. I also found that when the evaluators encountered critical incidents they (1) had experiences in their repertoire that helped them understand the situations they encountered; (2) utilized Recognition's Four By-Products; (3) sequentially generated decisions; and (4) used satisficing and an if-then-else decision making cycle to help them make decisions, similar to Klein's (1997) research on experts' decision making. When evaluators determined that their decision was not appropriate, needed to be modified, or lacked sufficient information, they asked stakeholders for clarification or conducted a new situation assessment.

Figure 7.1 depicts the findings that helped me understand expert evaluators' naturalistic decision making. I found that when conducting situation assessments and making decisions, evaluators relied on their philosophical assumptions (Mertens, 2009), interpersonal skills

(Johnson, 1998; Johnson, Lawrenz, & Volkov, 2009; Patton, 2009), self-determination (Baumeister, & Leary, 1995; Deci, & Ryan, 1985; McClelland, & Boyatis, 1980; Stevahn, King, Ghere, & Minnema, 2005), and satisficing (Simon, 1955). During data analysis, I found that evaluators' decision making consisted of three phases that led to a decision making cycle.

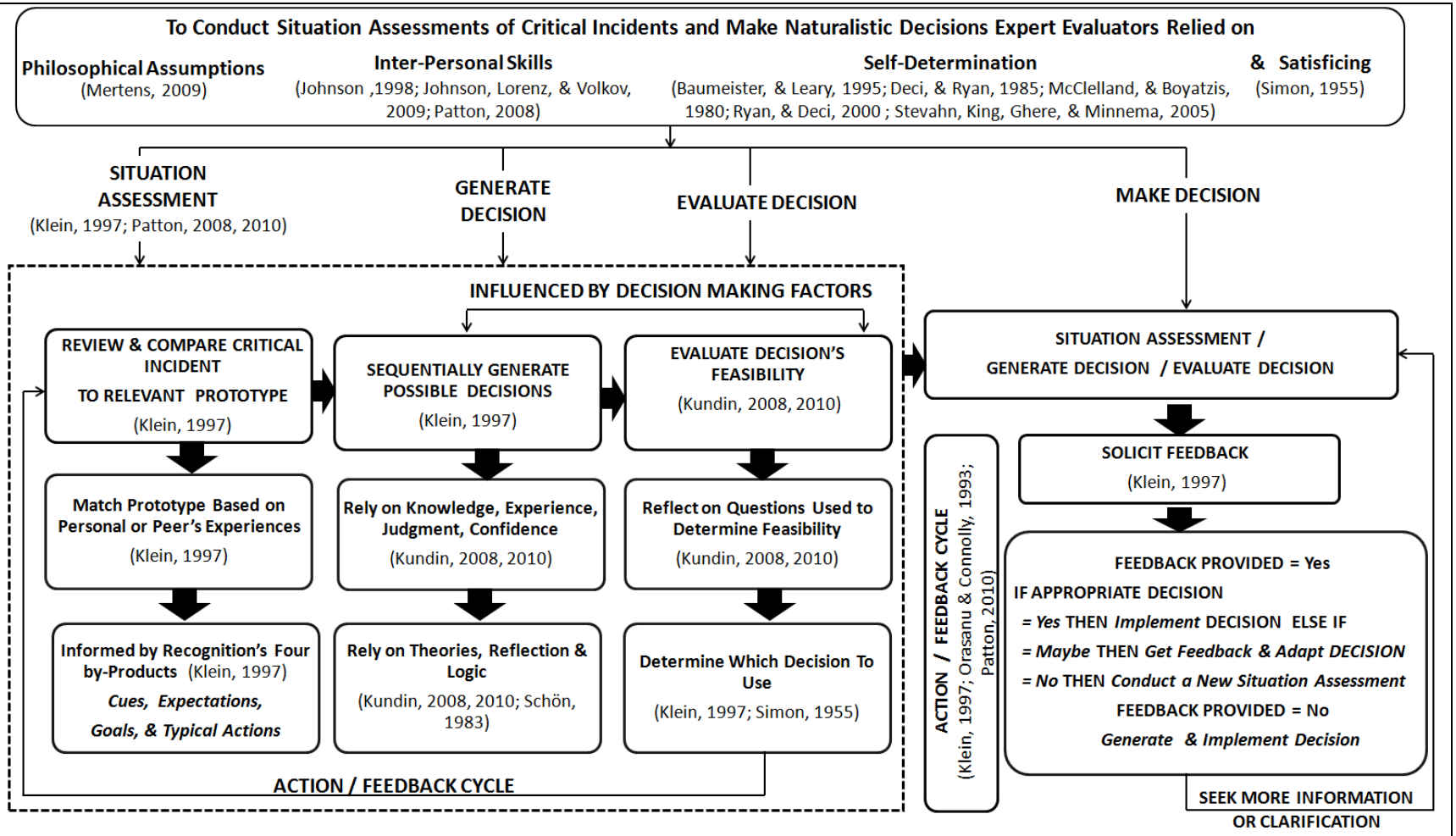

Figure 7.1. Expert evaluators' naturalistic decision making during critical incidents.

During the first phase, the evaluator conducted an initial situation assessment. As my study utilized critical incidents to understand the situations in which evaluators work, I found that evaluators conducted situation assessments by reviewing, comparing, and matching a critical incident to a relevant prototype. I found that evaluators relied on their personal experiences or peer-based experiences to identify a matching prototype. They matched the critical incidents to Recognition's Four by-Products. They compared the critical incidents encountered to what they expected to happen, utilized cues to determine what was occurring, compared their goals to their assumptions about stakeholders or other evaluators' goals, and compared the typical actions that would occur in a somewhat similar situation.

In my study, all of the evaluators were able to identify prototypes (Klein, 1997) that guided them when they generated a decision. During the second phase, the evaluators sequentially generated decisions by relying on their knowledge, experience, judgment, and confidence (Kundin, 2008, 2010). Evaluators also relied on theories (Kundin, 2008, 2010), reflection (Schön, 1983), and logic (Kundin, 2008, 2010) to help them make a probable decision. In the third phase, the evaluators evaluated a decision's feasibility. They reflected on whether a solution was feasible (Kundin, 2008, 2010). Based on their reflections and the previous cycles, the evaluators determined which decisions to employ. I also found that when evaluators were generating and evaluating their decisions, decision making factors influenced their naturalistic decision making. The three previous phases informed an evaluators' decision making cycle (see Figure 7.2).

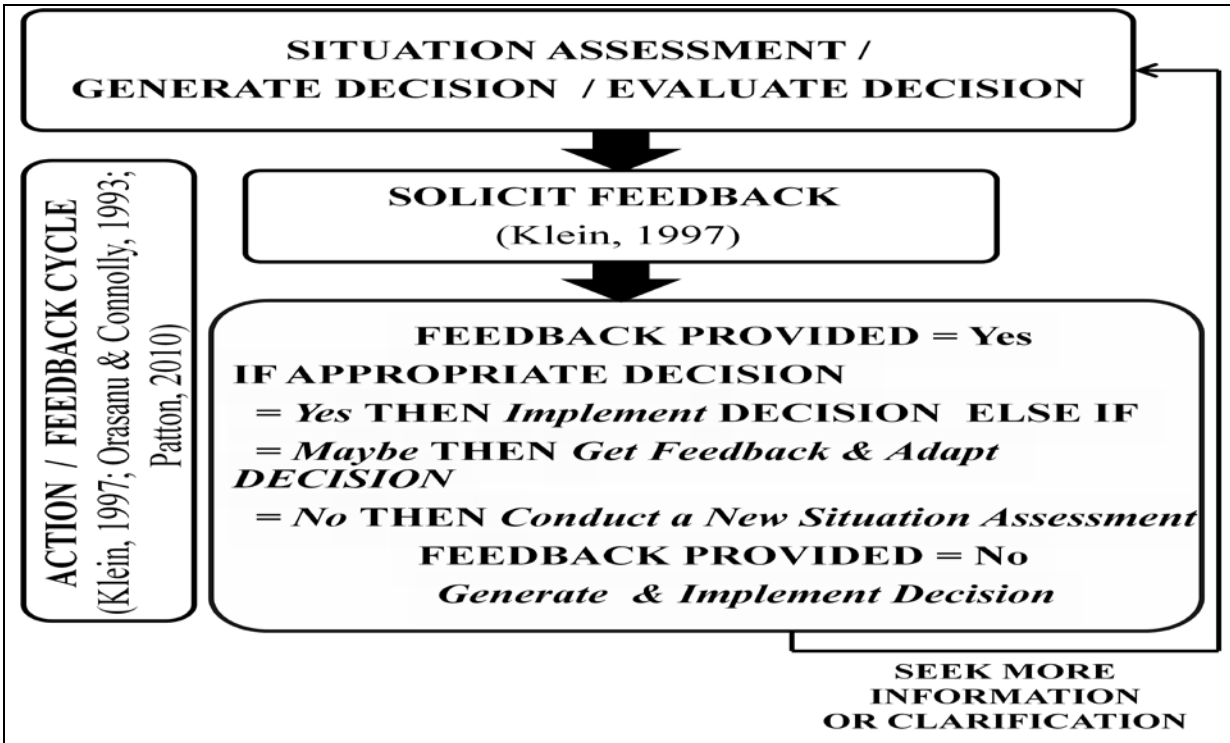

Figure 7.2. Expert Evaluator's Decision Making Cycle.

I found that the evaluators' decision making cycle is similar to Patton's (2008) decision making cycle: react, interact, and act. I also found that the evaluators used one or more action-feedback cycles to determine the appropriateness of a decision. If an evaluator determined that a decision was appropriate, he or she implemented the decision. However, if he or she determined that the decision might be appropriate, he or she modified the decision based on the feedback received. If the evaluator determined that the decision was not appropriate, he or she conducted a new situation assessment so that a new decision could be generated and implemented.

Nevertheless, my findings differ from Klein's (1997) PRD model. Klein's model confirmed that expert decision makers seek clarification or more information when their initial or subsequent decisions do not address the situation they have encountered. According to Klein, decision makers compare the probable decision or action to a prototype's typical decision or

action. My study found that the majority of evaluators conducted a complete situation assessment and used all of Recognition's Four by-Products. The evaluators re-evaluated the situation— based on their current understanding of a situation — by comparing the newly obtained information to a more informed prototype generated by matching and comparing the cues, goals, expectations, and typical actions. For example, in one case, an evaluator encountered evaluation resistance. When the evaluator made his initial decision, acted on his decision, and received feedback, he determined that his decisions and actions did not reduce evaluation resistance. Therefore, the evaluator looked at the cues (the evaluation participant was getting angrier), made an assumption about the evaluation participants' goal (avoiding evaluation because of the political environment, etc.), reviewed typical actions (based on a previous experience), and reviewed his expectations (evaluation will participate in an evaluation).

I also discovered that Klein's model does not address a circumstance in which a decision-maker did not receive feedback. None of the evaluators in my study stopped the evaluation because they had not received feedback. This indicated that the evaluators relied on Self-determination — their autonomy, knowledge, competence, relatedness, and self-regulation — to motivate themselves so that they could complete the evaluation, even though they did not receive the information they needed.

I found that whether or not expert evaluators received feedback, they relied heavily on their own assumptions (Mertens, 2009), experience, judgment, and confidence (Kundin, 2008, 2010) when making decisions. They also used reflection and logic to check their decisions in all phases of their decision making processes.

My study found that, like other experts, expert evaluators utilized satisficing because of the real-world constraints they encountered. My study also determined that the types of critical

incidents evaluators encountered were present in different stages of an evaluation's life cycle. Furthermore, the number of iterations through the action-feedback cycle was dependent on whether evaluators received feedback from stakeholders.

Differences in Critical Incidents, Decision Making Factors, and Decision Making

I utilized phenomenography to help me understand the qualitatively different ways that the evaluators described the following: the critical incidents they encountered, factors that influenced their decision making, and methods with which they made naturalistic decisions. Using a phenomenographic approach helped me generate a model that provided detailed information about evaluators' naturalistic decision making (see Figure 7.3). When I used the phenomenographic method I asked myself, "What are the different ways that evaluators are describing the critical incidents they encountered? How are evaluators describing the naturalistic decision making factors that influenced their decision making? Did the methods evaluators used to make naturalistic decisions vary?"

The key understandings found by using phenomenography indicated that the critical incidents were either barriers or enablers and expected or unexpected. The evaluators also reported that the factors influencing their naturalistic decision making were either positive or negative. An analysis of how evaluators made naturalistic decisions indicated that the number of decisions they made depended on whether they received stakeholder feedback.

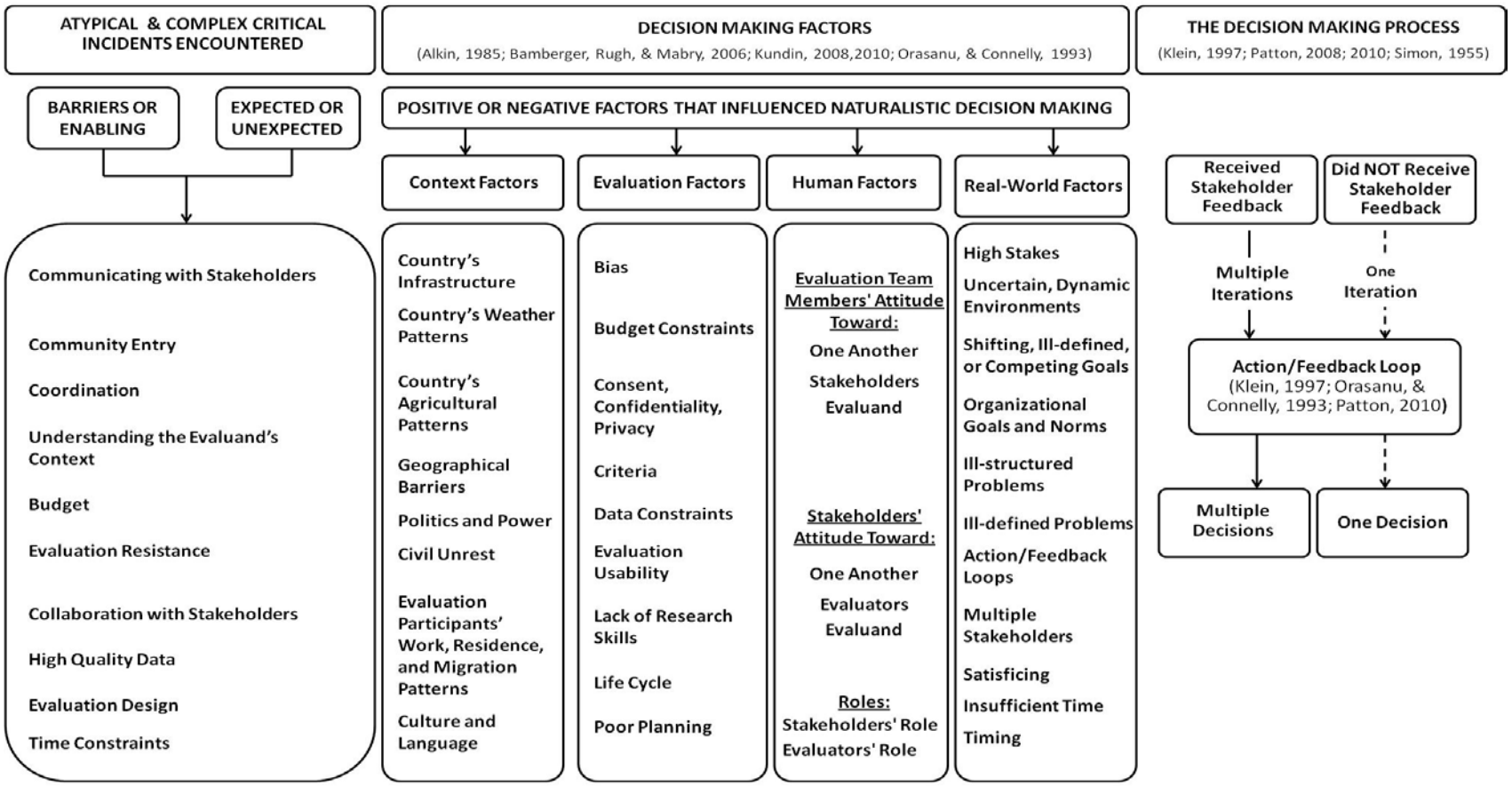

Figure 7.3. Variation among the critical incidents that the evaluators encountered, factors that influenced evaluators' naturalistic decision making, and the methods used when evaluators made decisions.

Critical incidents as barriers or enablers. A comparison of the narratives established that evaluators' efforts to address the critical incidents they encountered were not always successful. Evaluators reported critical incidents as barriers or enablers to planning, designing, and implementing the evaluation, and reporting the evaluation findings. For example, an evaluation team encountered critical incidents that were barriers to community entry. Likewise, an evaluation team experienced a successful collaboration with stakeholders that enabled the evaluation team to implement the evaluation. Further, an evaluator also stated that having sufficient time to complete an evaluation was helpful.

Expected or unexpected critical incidents. I found that evaluators reported that critical incidents were sometimes unexpected. For example, in Case One, an interviewee's hostility during an evaluation was unexpected. In another case, critical incidents were expected. For instance, in Case Seven, the evaluator stated that the evaluators anticipated that the evaluation instruments used in a different context would need to be adapted to the new evaluation community's context.

Negative or positive factors. My interpretation of the results indicated that evaluators reported decision making factors as positive or negative. For instance, the evaluator in Case One indicated that the stakeholder's negative attitude toward him and the evaluand influenced his decision to pursue the evaluation participant's participation in an interview. In another instance, evaluators found that stakeholders' positive attitude toward the evaluand had a positive influence on stakeholders' attitudes toward the evaluation.

Action-Feedback Loop. Evaluators indicated that the amount of decisions they made depended on whether they received stakeholder's feedback. If the evaluators received feedback, they performed multiple iterations of generating and seeking more information until a decision

was made and a course of action was determined. If the evaluators did not receive the requested stakeholder feedback, they made a single decision and did so in one iteration of the action/feedback loop.

Implications for Evaluation Practice

I found that evaluators addressed several issues when conducting evaluations. The issues addressed were as follows: Evaluation Resistance and Conflict, Consensus, Cooperation and Cooperation, Evaluation Use, Budget, Data and Time Constraints, Confidentiality, Privacy, and Consent, Evaluation Planning, Assumptions, Attitude and Evaluation Roles, and Silos of Evaluation Practice. This section provides recommendations based on the topics I deemed critical for evaluators to address.

Evaluation resistance and conflict. Evaluators encountered evaluation resistance in addition to conflict among and with stakeholders. Although evaluators cannot anticipate when evaluation resistance will happen, it is helpful to be aware that evaluation resistance is likely to occur at certain points during an evaluation's life cycle. Based on my findings, I recommend that evaluators prepare to deal with stakeholders who may feel threatened by an evaluation. I found that when evaluators were designing an evaluation, gathering data, and reporting the evaluation results the stakeholders felt threatened by the evaluation. Thus, awareness of what could be causing resistance to an evaluation may help evaluators address stakeholders' concerns.

The evaluators encountered evaluation resistance from over-evaluated evaluation participants. Consequently, I advise evaluators to determine how recently the programs have been evaluated. While this may not prevent evaluation resistance, evaluators will at least be aware that programs have recently been evaluated. I also propose that evaluators study the evaluand participants' population, agricultural, migratory, and work patterns so that their entry

into a community will not significantly interfere with the participants' activities. Avoiding interference with evaluation participants' routines can help evaluators forestall resistance to an evaluation or community entry.

My study also found that a stakeholder's perception of an evaluand influenced how he or she viewed an evaluation. Therefore, I recommend that the evaluators make efforts to understand how stakeholders perceive an intervention before the evaluation team conducts an evaluation. This could help them to understand or avoid evaluation resistance.

I found that evaluators used conflict management skills to handle conflict among evaluation partners during evaluation technical support meetings and when encountering evaluators' attempts to overcome evaluation resistance. The evaluators' use of conflict management skills points to the need for evaluators to learn how to manage conflict in evaluation settings (Stevahn, & King, 2005). It was observed that learning how to deal with conflict would assist evaluators as they make decisions because "evaluators almost inevitably experience conflict in the course of conducting evaluation studies" (p. 415).

Conflict is not limited to discord between stakeholders and evaluators. Evaluators also made decisions in situations involving civil unrest. This indicates that evaluators need to be aware of the social and political environment present when conducting an evaluation. Therefore, I advise evaluators to have an evaluation exit/evacuation plan because they perform development evaluations in dynamic and potentially volatile situations.

Conducting a stakeholder analysis can help evaluators identify key stakeholders, stakeholder roles, and their interests in the evaluation. I recommend that the evaluators conduct a stakeholder analysis so that they can create a strengths, weaknesses, and opportunities (SWOT) matrix. I also recommend that managers of evaluation teams use a SWOT matrix (Christensen,

Berg, & Salter, 1976) to determine their team's strengths, weaknesses, threats, and opportunities. A SWOT matrix will help to determine potential conflicts with and among stakeholders or evaluators.

Consensus, cooperation, and collaboration. The evaluators used literature reviews, structured observations, and stakeholder interviews to help them understand the evaluand's context prior to implementing an evaluation or securing community entry. The evaluators used logic models to build consensus among stakeholders so that the evaluators and stakeholders could work cooperatively and collaboratively. I recommend that evaluators be willing to adapt evaluation techniques to build consensus, cooperation, and collaboration between and among stakeholders so that the evaluation will proceed as smoothly as possible.

I also urge evaluators conducting cross-cultural evaluations to use participatory evaluation principles. Establishing partnerships and opportunities to collaborate with stakeholders will assist evaluators in implementing culturally sensitive evaluations. I also urge evaluators to establish partnerships so that the language used in the evaluation instruments is appropriate and yields usable data.

Evaluation use. In the evaluators' narratives, evaluators were concerned about the evaluation's use. The evaluators indicated that obtaining stakeholder feedback helped in designing a useful evaluation. It also helped them to determine how to make the evaluation useful and utilized by the agencies funding the evaluation.

The evaluators employed an intentional process use (Patton, 2008, 2010) — using the process of conducting an evaluation to build evaluation capacity — to ensure evaluation use. When planning an evaluation, the evaluators considered evaluation capacity building to be an important benefit, which should be provided to the evaluation stakeholders. Thus, it was

indicated that evaluators may need to assume the role of teacher so that they can provide stakeholders' evaluation capacity building. Therefore, I recommend training so that evaluators can deliver sound capacity-building opportunities for adult learners.

It was observed that the evaluators learned from the stakeholders' feedback. All of the evaluators' narratives highlighted the importance of communication and active listening. Consequently, I urge evaluators to develop excellent communication and listening skills.

Budget, data, and time constraints. I found that budgetary constraints and bureaucracy can result in poor quality evaluation data, affect an evaluation's project timeline (Bamberger et al, 2006), and require internal evaluators to secure supplemental financial resources for the evaluation. The evaluators' experiences revealed that budget constraints influenced the evaluators' ability to hire competent evaluation technical support. I urge evaluators to carefully budget for sufficient evaluation funds because they may not be able to implement a systematic and objective evaluation unless they have access to supplementary funds. I also found that evaluators' time constraints influenced the decisions they made about how the data would be gathered. Thus, evaluators were advised to plan for sufficient time to conduct the evaluation.

Privacy, confidentiality, and consent. Evaluators reported that it was necessary to secure the evaluation participants' consent. I found that different organizations and evaluators held different views regarding voluntary consent. Consequently, when conducting an evaluation, evaluators should be certain that they are aware of the evaluation funders' policy about the evaluation participants' required or voluntary consent. Similarly, funders should also be aware of the evaluators' views about evaluation participants' required or voluntary consent.

This indicates that when planning an evaluation, evaluators should include a plan describing how they would protect the research participants' privacy and confidentiality.

Furthermore, if evaluators are required to use an IRB protocol, they should be certain that they implement evaluations that are IRB compliant. Although evaluators are under pressure to gather data for the evaluation, they should ensure and certify that participants are not forced to participate in an evaluation and that there is informed consent.

Evaluation planning. The evaluators defined their type of evaluation as either formative or summative. It was proposed that when reading the evaluation requests for proposals or working with evaluation funders, learning when an evaluation will be conducted may guide the evaluator as he or she plans an evaluation. Further, knowledge regarding whether an evaluation is a formative or summative may help evaluators determine the evaluation's purpose, questions, and data that will inform the evaluation.

I recommend that evaluators use an evaluation-planning matrix that tracks the evaluation's purpose, evaluation criteria, evaluation question, evaluation data, and anticipated data gathering instruments. This may help evaluators determine whether the evaluation questions and data gathering activities are in line with the criteria used to judge the merit and worth of a program. Furthermore, it may help evaluators gather high quality data.

Assumptions, attitude and evaluation roles. My research found evidence that expert evaluators relied on their philosophical assumptions when making naturalistic decisions. This indicates that evaluators should reflect on their philosophical assumptions prior to accepting an evaluation job opportunity. Evaluators should also make their philosophical assumptions clear to key stakeholders. For example, if an evaluator considers social justice an important issue when conducting an evaluation, he or she should make it clear to the agency funding the evaluation. It is important to share such information because it can help the evaluation's funding agency determine if it is interested in hiring an evaluator with a particular point of view or approach.

It was observed that the stakeholder and evaluator roles in the evaluation influenced the evaluators' decision making. This indicates that evaluators should not limit their analysis to a stakeholder analysis, but should also be aware that their role in the evaluation and that of the stakeholders can influence how the evaluators conduct an evaluation and interact with stakeholders.

Similarly, stakeholders and evaluators' attitudes toward each other as well as the evaluand influenced evaluators' decision making. Although evaluators cannot control stakeholders' attitudes, they should reflect on their attitudes toward stakeholders or an evaluation prior to making a decision. Evaluator self-control was one of the factors that helped evaluators make naturalistic decisions. I recommend that evaluators learn how to deal with stressful situations since evaluators are working with stakeholders who have their own perceptions about the evaluator, the evaluation, and fellow stakeholders.

Silos of evaluation practice. Evaluators need to remember that implementing an evaluation in a real-world context does not always easily follow the procedures and theories taught in the classroom. As one evaluator stated, it is important to understand evaluation theory but real-world experience is essential if the evaluator wants to conduct rigorous evaluations. I recommend that evaluators share their evaluation experiences. Sharing evaluation stories will not only help spread evaluation knowledge among evaluators but also help practitioners or researchers from other disciplines to learn about evaluation.

Conclusions

To conclude, it can be said that the evaluators needed to modify or create new approaches when they encountered critical incidents. Further, the evaluators also needed to be innovative so that they could manage the characteristics present in real-world decision making environments.

In addition, the variety of critical incidents evaluators encountered, factors that influenced the evaluators' decision making, and the manner in which they made naturalistic decisions attests to the complexity inherent in each of the narratives. My findings also confirmed Patton's (2008) position that evaluations are complex. Although the evaluations were data driven, the evaluators' decision making was influenced by the context they were working in, evaluation factors, human factors, and characteristics found in real-world decision making.

The seven research participants were evaluators that had over ten years of evaluation experience and had graduate-level training in evaluation theory. Most of the research participants were affiliated with academic institutions or organizations where the research participants taught evaluation theory or provided training on how to conduct evaluations. The evaluators' narratives indicated that they relied on their knowledge on evaluation theory when planning, designing, and implementing an evaluation. Due to the atypical situations they encountered when conducting an evaluation, the evaluators also relied on skills—such as interpersonal skills—and previous evaluation experiences (personal or peer-based) to help them resolve conflict and unforeseen circumstances. The evaluators modified their evaluation plans, designs, and evaluation approaches to meet evaluation donor requirements, stakeholder's needs, and to obtain their goals, such as evaluation capacity building.

In summary, I found that evaluators encountered critical incidents when completing typical evaluation tasks, such as developing an evaluation design, gathering high quality evaluation data, and reporting evaluation findings. The critical incidents occurred because of budget, data, time constraints, and evaluation resistance. In addition, evaluators had to navigate onerous bureaucratic processes, understand the evaluand context, and secure community entry so that they could conduct the evaluation. I also found that evaluators needed to establish

partnerships to encourage collaboration and overcome language and cultural barriers. It was observed that the evaluators and stakeholders' perceptions of one another's roles, as well as their attitudes toward each other and the evaluand, influenced the evaluators' decision making. In addition, evaluators needed to use conflict management skills to overcome evaluation resistance caused by social and political pressures from external stakeholders.

Implications for Future Research

My study was aimed at understanding evaluators' decision making by studying the critical incidents that evaluators encountered and identifying the factors that influenced their decision making. However, further research is needed to understand the complex interaction between the critical incidents evaluators encountered, factors that influenced evaluators' decision making, and the manner in which evaluators made naturalistic decisions. I found the Critical Decision Method an effective method for understanding evaluators' naturalistic decision making. This method may be used to further understand the complex interactions between the critical incidents evaluators encountered, factors that influenced evaluators' naturalistic decision making, and the manner in which evaluators make naturalistic decisions.

Evaluations have common features, such as developing an evaluation design, selecting one or more methodological approaches, gathering data, analyzing data, and reporting evaluation findings. I found that each evaluation is unique, and therefore, evaluators must innovate or modify their evaluations in response to unexpected critical incidents or incidents that are barriers to conducting an evaluation. Studying evaluators as innovators (Patton, 2010) would help researchers to understand how people innovate in response to the critical incidents that they encounter.

My study focused on identifying the factors that influenced evaluators' decision making. The findings from a study of how decision making factors influence evaluators' decision making could teach evaluators how to deal with positive and negative factors that can influence their decision- making. Evaluators establish and maintain partnerships to encourage evaluation participation and collaboration (O'Sullivan, 2004). Conducting research on how expert evaluators established and maintained relationships with stakeholders could inform evaluation practice.

Chapter Summary

This chapter provided an overview of how I used the research design of my study. It also provided a summary of my study's findings, implications for practice and research, and conclusions. I studied the critical incidents that evaluators encountered, factors that influenced evaluator's naturalistic decision making, and the manner in which evaluators made their decisions, because I was interested in how the evaluators made naturalistic decisions during the evaluations of HIV/AIDS health education programs. I used a guiding framework, which was primarily based on Klein's (1997) and Kundin's (2008, 2010) research, to help me understand how evaluators made naturalistic decisions. The key findings were that evaluators encountered a variety of critical incidents that either were barriers to or enabled them in implementing evaluations and making decisions to address the critical incidents they encountered. Evaluators reported that they prepared for anticipated critical incidents but they also encountered unexpected events. The evaluators' naturalistic decision varied from Klein's (1997) naturalistic decision making model and from Kundin's (2008, 2010) Conceptual Framework for How Evaluators Make Practice Decisions. The findings obtained in my study will inform research on evaluators' decision making as well as naturalistic decision making.

REFERENCES

Alkin, M. (1985). *A guide for evaluation decision makers*. Thousand Oaks, CA: Sage.

Alkin, M. (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.

Alkin, M.C., & Christie, C.A. (2004). An evaluation theory tree. In M.C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 12–65). Thousand Oaks, CA: Sage.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Ashworth, P., & Lucas, U. (2000). Achieving empathy and engagement: A practical approach to the design, conduct and reporting of phenomenographic research. *Studies in Higher Education, 25*(3), 295–308.

Auerbach, J. D., Hayes, R. J. & Kandathil, S. M. (2006). Overview of effective and promising interventions to prevent HIV infection. In D. A. Ross, B. Dick, & J. Ferguson (Eds.), Preventing HIV/AIDS in young people: a systematic review of evidence from developing countries. UNAIDS Inter-Agency Task Team on Young People, Geneva, World Health Organization.

Baber, C., Fulthrope, C., & Houghton, R. J. (2010). Supporting naturalistic decision making through location-based photography: A study of simulated military reconnaissance. *International Journal of Human-Computer Interaction, 26*(3/4), 147–172.

Baker, D. P., Leon, J., & Collins, J. M. (2011). Facts, attitudes, and health reasoning about HIV and AIDS: Explaining the education effect on condom use among adults in Sub-Saharan Africa. *AIDS Behavior, 15*, 1319 –1327.

Bamberger, M. (2000). *Integrating quantitative and qualitative research in development projects.* Washington, DC: World Bank.

Bamberger, M. (2000). The evaluation of international development programs: A view from the front. *American Journal of Evaluation, 21*(1), 95–102.

Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation, 25*(1), 5–37.

Bamberger, M., Rugh, J., & Mabry, L. (2006). *RealWorld evaluation: Conducting evaluations with budget, time, data and political constraints.* Newbury Park, CA: SAGE.

Bandura, A. (1990). Perceived self-efficacy in the exercise of control over AIDS infection. *Evaluation and Program Planning, 13*, 9–17.

Banford, A., & Wong, B. L. W. (2004). Situation awareness in emergency medical dispatch. *International Journal of Human-Computer Studies, 61*, 421–452.

Barreto, M., & Ribeiro, S. L. (2012). Aircraft accident investigation: The decision-making in initial action scenario. *Work, 41*, 4931-4935.

Bartolo, P. A., Dockrell, J., & Lunt, I. (2001). Naturalistic decision-making task processes in multi-professional assessment of disability. *Journal of School Psychology, 39*(6), 499–519.

Baumeister, R., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497–529.

Baxter, G. D., Monk, A. F., Tan, K., Dear, P. R. F., & Newell, S. J. (2005). Using cognitive task analysis to facilitate the integration of decision support systems into the neonatal intensive care unit. *Artificial Intelligence in Medicine, 35*, 243–257.

Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts.* Chichester, UK: Wiley.

Becker, E., Jahn, T., Stiess, I., & Wehling, P. (1997). *Sustainability: A cross-disciplinary concept for social transformations* (Policy Papers No. 6). Paris: UNESCO.

Beeve, J. (2001). *Rapid assessment process: An introduction.* Walnut Creek, CA: Altamira Press.

Bickman, L., Heflinger, C. A., Pion, G., & Behar, L. (1992). Evaluation planning for an innovative children's mental health system. *Clinical Psychology Review, 12*, 853–865.

Blanford, A., & Wong, B. L. W. (2004). Situation awareness in emergency medical dispatch. *International Journal of Human-Computer Studies, 61*, 421–452.

Bogdan, R. C., & Bilken, S. K. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Boston: Allyn and Bacon.

Bogner, M. S. (1994). Human error in medicine: A frontier for change. In M. S. Bogner (Ed.) *Human error in medicine* (pp. 373–393). Hilldate, NJ: Lawrence Erlbaum Associates.

Bonoma, T. V. (1977) Overcoming resistance to changes recommended for operating programs. *Professional Psychology, 8*(4), 451–63.

Boyes, M. A., & O'Hare, D. (2011). Examining naturalistic decision making in outdoor adventure contexts by computer simulation. *Australian Journal of Outdoor Education, 15*(1), 22–34.

Bowden, J. (2005). Reflection on the phenomenographic research process. In J. Bowden & P. Green (Eds.), *Doing Developmental Phenomenology, Qualitative Research Methods Series,* (Vol 6.). Melbourne, Victoria: RMIT University Press.

Bowden, J., & Walsh, E. (Eds.). (2000). *Phenomenography.* Melbourne: RMIT University Press.

Boyle, & D. Lemaire (Eds.), *Building effective evaluation capacity*. New Brunswick, NJ: Transaction.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York, NY: Springer.

Breen, B. (2001, August 31). *What's your intuition?* Fast Company. Retrieved from http://www.fastcompany.com/magazine/38/klein.html?page=0%2C3

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18,* 32–42.

Bulmer, M. & Warwick, D. (1993). *Social research in developing countries: surveys and censuses in the third world*. London: Routledge.

Campbell, C. (2003). *Letting them die: Why HIV/AIDS intervention programs fail.* Cape Town, South Africa: International African Institute in association with Double Storey Books.

Campitelli, G. & Gobet, F. (2010). Herbert Simon's Decision-Making Approach: Investigation of cognitive processes in experts. *Review of General Psychology, 14(4),* 354–364.

Chavin, C., Clostermann, J., & Hoc, J. (2009). Impact of training programs in decision-making and situation awareness of trainee watch officers. *Safety Science, 47*(9), 1222–1231.

Chelimsky, E. (2001). What evaluation could do to support foundations: A framework with nine component parts. *American Journal of Evaluation, 22*, 13–28.

Chelimsky, E, & Shadish, W. R. (Eds.). (1997). *Evaluation for the 21st Century: A handbook.* Thousand Oaks, CA: Sage.

Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park: Sage.

Chen, H. T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283–302.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation, 30*(4), 457–494.

Christensen, C. R., Berg, N. A, & Salter, M. S. (1976). Policy formulation and administration: A casebook of top-management problems in business (7th edition). Homewood, ILL: R. D. Irwin.

Christie, C. A. (2009). Analyzing the practice of evaluation. In J. Fitzpatrick, C. A. Christie, & M. M. Mark (Eds.), *Evaluation in action: Interviews with expert evaluators* (pp. 393–435). Thousand Oaks, CA: Sage.

Christie, C. A., & Azzam, T. (2005). What theorists say they do: A brief description of theorists' approaches. *New Directions for Evaluation, 106,* 15–26.

Christie, C. A & Rose, M. (2003). Learning about evaluation through dialogue: Lessons from an informal discussion group. *American Journal of Evaluation, 24*(2), 235–243.

Compton, D., Baizerman, M., & Stockdill, S.H. (Eds.). (2002). *The art, craft, and science of evaluation capacity building*. San Francisco, CA: Jossey-Bass.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: A technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science, 16*(1), 41–51.

Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* Cambridge, MA: MIT Press.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum

Denzin, N. K., & Lincoln, Y. S., (2008). *Collecting and Interpreting Qualitative Materials* (3rd ed.). Thousand Oaks, CA: Sage.

Dewey, J. D., Montrosse, B. E., Schröter, D. C., Sullins, C. D., & Mattox III, J. R. (2008, Sept.). Evaluator competencies : What's taught versus what's sought. *American Journal of Evaluation, 29*(3), 268–287.

Di Bello, L. (1997). Exploring the relationship between activity and expertise: Paradigm shifts and decision defaults among workers learning materials requirements planning. In C. E. Zsambok & G. Klein (Eds.) *Naturalistic decision making* (pp. 163–174). Mahwah, NJ: Lawrence Erlbaum Associates.

Dillon, S., Buchman, J., & Corner, J. (2005). A proposed framework of descriptive decision making theories. *Icfaian Journal of Management Research, 5*(2), 65–74.

Donaldson, S. I., Christi, C. I., & Mark, M. M. (2008). *What counts as credible evidence?* Thousand Oaks, CA: Sage.

Earl, S., & Carden, F. (2002). Learning from complexity: The international development research centre's experience with outcome mapping. *Development in Practice, 12*(3 & 4), 518–524.

Eastman, C. M. (1969) Cognitive processes and ill-defined problems: a case study from design. In: Walker DE, Norton LM (Eds.) *Proceedings of the international joint conference on artificial intelligence (IJCAI '69),* Washington, DC, May 1969, pp 675–699.

Ebbutt, D. (1998). Evaluation of projects in the developing world: Some cultural and methodological issues. *International Journal of Educational Development, 18*(10), 415–424.

Edwards, W. (1954). The theory of decision making, *Psychological Bulletin, 51*, 380–417.

Elkins, C. (2010). Evaluating Development Interventions in Peace-Precarious Situations. E*valuation,* 16(3), 309–321.

Elliot, J. (2005). *Using narrative in social research: Qualitative and quantitative approaches.* London: Sage.

Endsley, M. R. & Garland, D. J. (2000). *Situation awareness analysis and measurement.* Mahwah, N.J: Lawrence Erlbaum Associates.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363–406.

Ericsson, K. A., & Simon, H. A. (1980). *Verbal reports as data. Psychology Review, 87(3),* 215–251.

Fitzpatrick, J. L. (2004). Exemplars. *American Journal of Evaluation, 25*(4), 541–559.

Fitzpatrick, J. L., Christie, C., & Mark, M. M. (2009). *Evaluation in action: Interviews with expert evaluators*. Thousand Oaks, CA: Sage.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. J. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4rd ed.). Upper Saddle River, NJ: Pearson.

Fisher, J. D. (1992). Changing AIDS – Risk Behavior: Center for Health, Intervention, and Prevention (CHIP). *Psychological Bulletin, 111*(3), 455–474.

Flin, R., Stewart, K., & Slaven, G. (1996). Emergency decision making in the offshore oil and gas industry. *Human Factors, 38*(2), 1–15.

Franklin, A., Liu, Y. N., Johnson, T. R., Robinson, D., Okafor, N., King, B., . . . Zhang, J. (2011). Opportunistic decision making and complexity in emergency care. *Journal of Biomedical Informatics*(3), 469–476.

Freeman, J., Weitzenfeld, J., Klein, G., Ried, T., & Musa, J. (1991, April). *A knowledge elicitation technique for education development: The critical decision method*. Paper presented at the Annual Conference of the American Educational Research Association, Chicago, IL.

Friedman, R. M. (1997). *Evaluating mental health services: How do programs for children "work" in the real world?* Thousand Oaks, CA: Sage.

Frierson, H. T., Hood, S. & Hughes, G. B. (2002). A guide to conducting culturally responsive evaluations. In J. Frechtling, The 2010 user-friendly handbook for project evaluation (pp. 75-96). Arlington, VA: National Science Foundation.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*(4), 331–362.

Gajda, R. (2004). Utilizing collaboration theory to evaluate strategic alliances. *American Journal of Evaluation, 25(*1), 65–77.

Glesne, C., & Peshkin, A. (1992). *Becoming Qualitative Researchers: An Introduction.* White Plains, NY: Longman.

Greene, J. C. (2005a). Context. In S. Matheson (ED.) *Encyclopedia of evaluation* (pp. 82—84). Thousand Oaks, CA: Sage Publications.

Greitzer, F. L., Podmore, R., Robinson, M., & Ey, P. (2010). Naturalistic decision making for power system operators. *International Journal of Human-Computer Interaction, 26*(2/3), 278–291.

Handwerker, W. P. (2001). *Quick ethnography*. Walnut Creek, CA: Altamira Press.

Helslooot, I., & Groenendall, J. (2011). Naturalistic decision making in forensic science: Toward a better understanding of decision making by forensic team leaders. *Journal of Forensic Sciences, 56*(4), 890–897.

Hernandez, M. (2000). Using logic models and program theory to build outcome accountability. *Education and Treatment of Children, 23(1),* 24–40.

Hernandez, M., Hodges, S. P., & Cascardi, M. (1998). The ecology of outcomes: System accountability in children's mental health. *The Journal of Behavioral Health Services and Research*, *25*(2), 136–150.

Hernandez-Serrano, J. (2001). Effects, meaning, and learning processes regarding expert's stories and novice problem solving in an ill-structured problem solving environment (Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3016659)

Hobbs, J. R. (2003). The framework: Interpretation as abduction. In J. Hobbs, *Discourse and inference: Magnum Opus in Progress* (Chapter 3). Retrieved from http://www.isi.edu/~hobbs/disinf-tc.html

Hoffman, R. R. & Militello, L. G. (2009). *Perspectives on cognitive task analysis: Historical origins and modern communities of practice.* New York, NY: Psychology Press.

Hogard, E., & Ellis, R. (2006). Evaluation and communication: Using a communication audit to evaluate organisational communication. *Evaluation Review*, *30*(2), 171–187.

Hood, S. (2004). A journey to understand the role of culture in evaluation: Snapshots and personal reflections on one African American evaluator. In M. Thompson-Robinson, R. Hopson, & S. SenGupta (Eds.), *In search of cultural competence in evaluation: Towards*

principles and practices. New Directions for Evaluation (No. 102, pp. 21–37). San Francisco, CA: Josey-Bass.

Hood, S., Hopson, R., & Frierson, H. (Eds.) (2005). *The role of culture and cultural context in evaluation: A mandate for inclusion, the discovery of truth, and understanding in evaluation theory and practice.* Charlotte, NC: Information Age.

House, E. R., Haug, C., & Norris, N. (1996). Producing evaluations in a large bureaucracy. *Evaluation*. *2*(2), 135–150.

Hutchins, S. G. (1997). Decision-making errors demonstrated by experienced naval officers in a littoral environment. In C. E. Zsambok & G. Klein (Eds.) *Naturalistic decision making* (pp. 207–215). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hutton, R. J. & Klein, G. (1999). Expert decision making. *Systems Engineering, 2*(1), 32–45.

Ireland, J., Tambyah, M., Neofa, Z., & Harding, T. (2008). The tale of four researchers: Trials and triumphs from the phenomenographic research specialization. Paper presented at the 2008 Australian Association for Research in Education International Education Research Conference (AARE), November 4th-December 4th, 2008. Brisbane, Australia.

Jenkins, D. P., Stanton, N. A., Salmon, P. M., Guy, H., & Rafferty, L. (2010). Using the decision-ladder to add a formative element to naturalistic decision making research. *Journal of Human-Computer Interaction*, *26*(2/3), 132–146.

Johnson, Jr., B. L., & Fauske, J. R. (2000). Principals and the political economy of environmental enactment. *Educational Administration Quarterly, 36,* 159–185.

Johnson, K., Greenseid, L., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 – 2005. *American Journal of Evaluation, 30*(3), 377–410.

Johnson, R. B. (1998). Toward a theoretical model of evaluation. *Evaluation and Program Planning, 21*(1), 93–110.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards.* Thousand Oaks, CA: Corwin Press.

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology: Research and Development, 45*(1), 65–94.

Julian, D. A. (1997). The utilization of the logic model as a system level planning and evaluation device. *Evaluation and Program Planning, 20*(3), 251–257.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Kaempf, G. L., & Orasanu, J. (1997). Current and future applications of naturalistic decision making in aviation. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 305–318). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kahneman, D., Slovic, P., and Tversky A. (Eds.). (1982*). Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kirby, D., Obasi, A., Larris, B. A. (2006). The effectiveness of sex education and HIV education interventions in schools and developing countries. In D. Ross, et al, *Preventing HIV/AIDs in Young People: A Systematic Review of the Evidence from Developing Countries (pp. 103-150),[WHO Technical Report Series, 938]*, Geneva: WHO and Inter-Agency Task Team on HIV and Young People.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation, 31*(3), 400–413.

Klein, G., & Brezovic, C. P. (1986). Design engineers and the design process: Decision strategies and human factors literature. In *Proceedings of the 30th Annual Meeting of the Human Factors Society* (pp. 771–775). Santa Monica, CA: Human Factors Society.

Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods,* (pp. 138–147). Norwood, NJ: Ablex.

Klein. G. A. (1998). *Sources of Power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. (2008). Naturalistic Decision Making. Human factors: *The Journal of Human Factors and Ergonomics Society, 50*(3), 456–460.

Klinger, D. W., Andriole, S. J., Militello, L. G., Adelman, L., Klein, G., & Gomes, M. E. (1993). *Designing for performance: A cognitive systems engineering approach to modify an AWACS human-computer interface* (AL/CF-TR-1993-0093). Wright-Patterson AFB, OH: Department of the Air Force, Armstrong Laboratory, Air Force Material Command.

Kniker, T. (2011). Evaluation survivor: How to outwit, outplay and outlast an internal government evaluator. In B. B. Volkov & M. E. Baron (Eds.). *Internal Evaluation in the 21st Century: New Directions for Evaluation, 132*, 57–72.

Krippendorff, K. (2004). *Content Analysis: An introduction to its methodology.* Thousand Oaks, CA: Sage.

Kundin, D. M. (2008). *Everyday approaches to evaluation: A study of how evaluators make practice decisions.* (Doctoral dissertation). Retrieved from ProQuest. AAT 3299404

Kundin, D. M. (2010, June). A conceptual framework for how evaluators make everyday decisions. *American Journal of Evaluation, 31*(3), 347–362.

Kusek, J. Z., & Rist, R. C. (2004). Ten steps to a results-based monitoring and evaluation system. Washington, DC: The World Bank.

Kvale, S. (2007). Doing interviews. In *The Sage Qualitative Research Kit*, Uwe Flick (Ed.), Thousand Oaks, CA: Sage.

Lave, J. (1998*). Cognition in practice: Mind, mathematics, and culture in everyday life.* Cambridge: Cambridge University Press.

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Lipshitz, R. (1987). Decision making in the real world: Developing descriptions and prescriptions from decision maker's retrospective accounts, Report, *Center for Applied Science*. Boston: Boston University.

Lipshitz, R., & Shaul, O. B. (1997) Schemata and mental models in recognition-primed decision making. In C. E. Zsambok & G. Klein (Eds.) *Naturalistic decision making* (pp. 293–303). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2000). *Taking stock of naturalistic decision making.* Retrieved from http://organizations.haifa.ac.il/html/html_eng/raanan%20-%20taking.doc.

Littig, B. (2007, May). *Interviewing experts or interviewing the elite: What makes the difference?* Paper presented at the Conference on Interpretation in Policy Analysis, Amsterdam, Netherlands. Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1000

Locke, E. A., & Latham, G. P. (1990). *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice-Hall.

Ludema, J. D. & Virgilio, M. E. D. (2007). The role of energy in conversation in leading organizational change. In W. A. Passmore & R. W. Woodman (Eds.), *Research in organization change and development*, (Vol. 16, pp. 1–42). Greenwich, CT: JAI Press.

Manswell, B., J., Reid, M., and LaPoint, V. (2004). A culturally responsive evaluation approach applied to the Talent Development School-to-Career Intervention Program. *New Directions for Evaluation, 101*, 37–47.

Mantell, J. E., DiVittis, A. T., & Auerbach, M. I. (1997). *Evaluating HIV Prevention Interventions.* New York: Plenum.

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

Marshall, A. P., West, S. H., & Aitken, L. M. (2011). Preferred information sources for clinical decision making: Critical care nurses perceptions f information accessibility and usefulness. *Worldviews on Evidence-based Nursing*, *8*(4), 224–235.

Marton, F. (1988). Phenomenography: A research approach to investigating different understandings of reality. In R. R. Sherman & R. B Webb (Eds.), *Qualitative research in education: Focus and methods* (pp. 141–161). New York: Falmer Press.

Marton, F. (1994). Phenomenography. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed., Vol. 8, pp. 4424–4429). Oxford, UK: Pergamon.

Marton, F., & Booth, S. (1997). *Learning and Awareness*. Mahwah, NJ: Lawrence Erlbaum Associates.

Mattessich, P. W., Murray-Close, M., & Monsey, B. R. (2001). *Collaboration: What makes it work (*2nd ed.). St. Paul, Minnesota: Amherst H. Wilder Foundation.

McClelland, D. C., & Boyatzis, R. E. (1980). Opportunities for counselors from the competency assessment movement. *Personnel and Guidance Journal*, *58*(5), 368–372.

McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning*, *22*(1), 65–72.

MEASURE. (2007). *Data quality assurance tool for program-level indicators* [PEPFAR and U.S. Agency for International Development (USAID) Report no. MS-07-19]. Chapel Hill, NC: University of North Carolina.

Mendonca, D. (2007). Decision support for improvisation in response to extreme events: Learning from the response to the 2001 World Trade Center attack. *Decision Support Systems*, *43*, 952–967.

Merriam, S. B. (1998). *Qualitative research and case study applications in education (Rev. ed.).* San Francisco: Jossey-Bass Publishers.

Mertens, D. M. (2008). *Transformative research and evaluation.* New York, NY: The Guilford Press.

Mertens, D. M., & Wilson, A. T. (2012). *Program Evaluation Theory and Practice: A Comprehensive Guide.* New York: The Guilford Press. Retrieved from Google Books Preview (Pending title release).

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis (2nd ed.) Thousand Oaks, CA: Sage.

Miller, T. E., Pyle, D. M., & Shore, J. S. (1993). Development of a prototype decision support system for munitions effects assessment. *Proceedings of the special session held concurrently with the Sixth International symposium in Interaction of Nonnuclear*

Munitions with Structures (pp. 1–2). Vicksburg, MS: U. S. Army Engineer Waterways Experiment Station.

Mitchell, C. M., Morris, J. G., Ockerman, J. J., & Potter, W. J. (1997). Recognition primed decision making as a technique to support reuse in software design. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 305–318). Mahwah, NJ: Lawrence Erlbaum Associates.

Mooney, J. (1947). *The principles of organization*. New York: Harper & Brothers.

Morell, J. A. (2010). *Evaluation in the face of uncertainty: Anticipating surprise and responding to the inevitable.* New York: The Gilford Press.

Morra Imas, L. G., & Rist, R. C. (2009). *The road to results: Designing and conducting effective development evaluations*. Washington, D. C.: World Bank Publications.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315–340.

National Joint Committee for the Communication Needs of Persons With Severe Disabilities. (1992). *Guidelines for meeting the communication needs of persons with severe disabilities* [Guidelines]. Available from http://www.asha.org/docs/pdf/GL1992-00201.pdf

Newton, P. M., & Sackney, L. (2005). Group knowledge and group knowledge processes in school board decision making. *Canadian Journal of Education*, *28*(3), 434–457.

Nikolaev, A. G. (2007). *International negotiations: Theory, practice, and the connection with domestic politics*. Lanham, MD: The Rowman & Littlefield Publishing Group.

Nyanzi, S. (2006). From miniscule biomedical models to sexuality's depths. *Lancet, 368*(9550), 1851-1852.

OECD (Organisation for Economic Co-operation and Development) (1991). *Principles for evaluation of development assistance. Development Assistance Committee*. Paris: OECD Publishing.

OECD (Organisation for Economic Co-operation and Development) (2008) *Paris Declaration on Aid Effectiveness and Accra Agenda for Action. Paris:* OECD Publishing.

OECD. (2010). *Quality standards for development evaluation*. DAC Guidelines and Reference Series. Paris: OECD Publishing.

Okonta, P. I., & Oseji, M. I. (2006). Relationship between knowledge of HIV/AIDS and sexual behaviour among in-school adolescents in Delta State, Nigeria. *Nigerian Journal of Clinical Practice, 9*(1), 37–39.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A Klein, J. Orasanu, R. Calderwood, & E. E. Zsambok (Eds.*), Decision making in action: Models and Methods* (pp. 3–20). Norwood, NJ: Ablex.

Owen, J. M. (2007). *Program evaluation: Forms and approaches***.** New York: Guilford Press.

Parlett, M., & Hamilton, D. (1976). Evaluation as illumination? A new approach to the study of innovative programmes. In G. Glass (Ed.). *Evaluation Studies Review Annual 1*, 140–157.

Pascual, R., & Henderson, S. (1997). Evidence of naturalistic decision making in military command and control. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 217–226). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Patton, M. Q. (2008). Utilization-Focused evaluation. Thousand Oaks, CA: Sage.

Patton, M. Q. (2002). *Qualitative evaluation and research methods* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* Boston, MA: Guildford Press.

Peshkin, A. (1988). In search of subjectivity—one's own. *Educational Researcher, 17*(7), 17–21.

Preskill, H. (1997). Using critical incidents to model effective evaluation practice in the teaching of evaluation. *American Journal of Evaluation, 18*(1), 65–71.

Reeves, T. C., & Hedberg, J. G. (2003). *Interactive learning systems evaluation*. Englewood Cliffs, NJ: Educational Technology Publications.

Regeer, B.J., Hoes, A., van Amstel-van Saane, M., Caron-Flinterman, F.F., & Bunders, J.F.G. (2009). Six Guiding Principles for Evaluating Mode-2 Strategies for Sustainable Development. *American Journal of Evaluation, 30*(4), 515–537.

Research Technology International (RTI). (2004). *Community-based participatory research: Assessing the evidence.* (Report/Technology Assessment no. 99). Rockville, MD: Agency for Healthcare Research and Quality..

Resnick. L. B., Levine, J. M., & Teasley, S. D. (Eds.). (1996). *Perspectives on socially shared cognition.* Washington, DC: American Psychological Association.

Ross, D. A., Wight, D., Dowsett, G., Buve, A., & Obasi, A. (2006). The weight of evidence: A method for assessing the strength of evidence on the effectiveness of HIV prevention interventions among young people. *World Health Organization Technical Report Series, 938*, 79–102.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach.* Thousand Oaks, CA: Sage.

Roth, E. M. (1997). Analysis of decision making in nuclear power emergencies: an investigation of added decision making. In G. A Klein, J. Orasanu, R. Calderwood, & E. E. Zsambok (Eds.*), Decision making in action: Models and methods* (pp. 3–20). Norwood, NJ: Ablex.

Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist, 45*(4), 489–493.

Running Wolf, P., Sloer, R., Manteuffel, B., Sondheimer, D., Santiago, R., & Erickson, J.S. (2002). Cultural competence approaches to evaluation in tribal communities. In J.D. Davis, J.S. Erickson, S.R. Johnson, C.A. Marshall, P. Running Wolf, & R. Santiago (Eds.). *Workgroup on American Indian research and program evaluation methodology (AIRPEM), Symposium on research and evaluation methodology: Lifespan issues related to American Indians/Alaska Natives with disabilities*, pp. 32–49. Flagstaff, AZ.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Salabarria-Peña, Y., Apt, B. S., & Walsh, C. M. (2007). *Practical use of program evaluation among sexually transmitted disease (STD) programs.* Atlanta, GA: Centers for Disease Control and Prevention.

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (2009). How can you turn a team of experts into an expert team? Emerging training strategies. In C. E. Zsambok & G. A. Klein (Eds.), *Naturalistic decision making* (pp. 359–370), New York, NY: Routledge.

Saldana, J. (2009). *The coding manual for qualitative researchers.* Thousand Oaks, CA: Sage.

Sanders, J. R. (2002). A vision of evaluation. *American Journal of Evaluation, 23*(3), 253–259.

Sattler, J. L. & Armmer, F. (2006). Inquiry-based approaches to HIV/AIDS prevention education: Standards, practices and outcomes. In American Association of Colleges for Teacher Education with Carl A. Grant and Liane M. Summerfield in *Humanizing Pedagogy through HIV and AIDS Prevention: Transforming Tteacher Knowledge* (pp. 75–96). Boulder, CO: Paradigm.

Schank, R. C. (1990). *Tell me a story: Narrative and intelligence*. Evanston, IL: Northwestern University Press.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schraagen, J. M. (1997). Discovering requirements for naval damage control decision support systems. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 227–232). Mahwah, NJ: Lawrence Erlbaum Associates.

Schwandt, T., & Dahler-Larson, P. (2006). When evaluation meets the 'Rough Ground' in communities. *Evaluation, 12*, 496–505.

Schwarz, C., & Stuhkamp, G. (2007). Does evaluation build or destroy trust? Insights from case studies in evaluation in higher education reform. *Evaluation, 13*(3), 323-339.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology, 83*(5), 1178–1197.

Scrimshaw, S., & Hurtado, E. (1987). *RAP: Rapid assessment procedures for nutrition and primary health care*. Tokyo: United Nations University.

Scriven, M. (1967). The methodology of evaluation. In R. F. Stake (Ed.), *Curriculum Evaluation,* (American Educational Research Association Monograph Series on Evaluation, No. 1, pp. 39-83). Chicago; Rand McNally.

Serfaty, D., MacMillan, J., Entin, E. E., & Entin, E. B. (1997). The decision making expertise of battle commanders. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 233–246). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Simon, H. A. (1947). *Administrative behavior*: *A study of decision making processes in administrative organization*. New York: The Free Press.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence, 4*(3–4), 181–201.

Simon, H. (1947). *Administrative behavior*: *A study of decision making processes in administrative organizations*. New York: The Free Press.

Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science, 2*(1), 125–134.

Smith, G. F. (1997). Managerial problem solving: A problem-centered approach. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 371–380). Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, T. (1990) Policy evaluation in third world countries: Some issues and problems. *The Asian Journal of Public Administration, 12,* 55–68.

Solomon, B. (2002). Accountability in public child welfare: Linking program theory, program specification and program evaluation. *Children and Youth Services Review, 24,* 385–407.

Sonnichsen, R. C. (2000). *High impact internal evaluation: A practitioner's guide to evaluating and consulting inside organizations.* Thousand Oaks, CA: Sage.

Spicer, N., & Walsh, A. (2011). 10 best resources on . . . the current effects of global health initiatives on country health systems. *Health Policy and Planning,* 27(3), 265–269.

Spooner, C., Flaxman, S., & Murray, C. (2008). Participatory research in challenging circumstances: Lessons with a rural aboriginal program. *Evaluation Journal of Australasia, 8*(2), 28–34.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation. 26*(1), 43–59.

Stinchcomb, J. B. (2001). Using logic modeling to focus evaluation efforts: Translating operational theories into practical measures. *Journal of Offender Rehabilitation, 33,* 47–66.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications* (1st ed.). San Francisco: Jossey-Bass.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication.* New York: Cambridge University Press.

Svenson, O., & Maule, A. J. (1993). *Time pressure and stress in human judgment and decision making*. New York: Plenum.

Taut, S. & Brauns, D. (2003). Resistance to evaluation: A psychological perspective. *Evaluation, 9*(3), 247–264.

Tourmen, C. (2009). Evaluators' decision making: The relationship between theory, practice, and experience. *American Journal of Evaluation, 30*(1), 7–30.

Theroux, R. (2010). Women's decision making during the menopausal transition. *Journal of the American Academy of Nurse Practitioners, 22*(11), 612-621.

UNAIDS. (2008). *Report on the global HIV/AIDS epidemic 2008: Executive summary*. Geneva, Switzerland: UNAIDS.

UNAIDS. (2011). *World AIDS Day Report 2011: How to get to zero- Faster. Smarter. Better., United Nations Programme on HIV/AIDS.* Geneva, Switzerland: UNAIDS.

UNDP (United Nations Development Programme) Evaluation Department. (1998) *The programme approach: Ownership, partnership and coordination*. New York: UNDP.

U. S. Office of Management and Budget. (2009). *A new era of responsibility: Renewing America's Promise.* New York, NY: Cosimo Reports.

Vamjaramta, M., & Easton, G. (2010). Intuitive managerial thinking: the use of mental simulations in the industrial marketing context. *Industrial Marketing Management, 39*(3), 425–436.

VanWynsberghe, R. (2007). Redefining case study. *International Journal of Qualitative Methods*. *6*(2), 1–10.

Visser, W. (2000, May). A tribute to Simon and some - too late – questions by a cognitive ergonomist. INRA Report 4462. Rocquencourt, France: INRIA

Volkov, B. B. (2011). Beyond being an evaluator: The multiplicity of roles of the internal evaluator. In B. B. Volkov & M. F. Baron (Eds.) *Internal evaluation in the 21st century: New directions for evaluation, 132,* 25–42.

Volkov, B. B., & Baron, M. E. (2011). Issues in internal evaluation: Implications for practice, training, and research. In B. B. Volkov & M. F. Baron (Eds.) *Internal evaluation in the 21st century: New directions for evaluation, 132,* 25–42.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Waag, W. L., & Bell, H. H. (1997). Situation assessment and decision making in skilled fighter pilots. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 247–254). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Walker, R., Hoggart, L., & Hamilton, G. (2008). Random assignment and informed consent: A case study of multiple perspectives. *American Journal of Evaluation, 29*(2), 156–174.

Wang, K. C., Hsieh, A. T., & Huan, T. C. (2000). Critical service features in group package tours: An exploratory research. *Tourism Management, 21*, 177–189.

Watzlawick, P., Weakland, J.W. and Fisch, R. (1974). *Change*. New York: Norton.

Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review, 21*(4), 501–524.

Weiss, C. H. (1998) *Evaluation* (2nd ed). Upper Saddle River, NJ: Prentice Hall.

Weiss, D. H. (1990). *Motivoi alaisesi huippusuorituksiin (Motivate your subordinates to top performance).* Tampere: Mäntän Kirjapaino Oy.

Weldon, M. S. (2001). Remembering as a social process. *Psychology of Learning and Motivation, 40*, 67–120.

Wisler, C. (Ed.). (1996). Evaluation and auditing: Prospects for convergences, *New Directions for Evaluation, 71,* 1–71. San Franciso: Jossey-Bass.

Wolcott, H. F. (1995). *The art of fieldwork.* Walnut Creek, CA: AltaMira Press.

World Bank. (2006) *Conducting quality impact evaluations under budget, time, and data constraints.* Washington, DC: Independent Evaluation Group.

World Bank. (2007). Evaluating the World Bank's Assistance for Fighting the HIV/AIDS Epidemic. Retrieved from http://www.worldbank.org/ieg/aids/case_studies.html

World Health Organization. (2002). *Global summary of the AIDS epidemic.* World Health Organization. Retrieved from http://www.who.int/hiv/data/2011_epi_core_en.png

Wortham, S. (2001). Interactionally situation cognition. A classroom example. *Cognitive Science, 25,* 37–66.

Wottawa, H., & Thierau, H. (1998) *Lehrbuch Evaluation* (2nd ed). Bern: Hans Huber.

Xiao, Y., Milgram, P., & Doyle, D. J. (1997. Capturing and modeling planning expertise in anesthesiology: Results of a field study. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 197–205). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yin, R. K. (1995). *Case study research: Design and method* (2nd ed). Thousand Oaks, California: Sage.

Yin, R. K. (2003). *Case study research, design and methods* (3rd ed). Newbury Park: Sage.

Zsambok, C. E., & Klein, G. (1997). *Naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

APPENDIX A

PILOT STUDY

Program Evaluator's Decision Making When Facing Critical Incidents

Abstract: Evaluators' decision making has an impact on program evaluation planning, design implementation, and reporting. This pilot study explores a program evaluator's decision making when facing critical incidents. Using the critical decision method, a narrative was generated and analyzed. The narrative provided an example of the critical incidents that the evaluator faced and revealed aspects of the evaluator's naturalistic decision making. A decision making model was generated to reveal the evaluator's decision making process. Potential applications of this research include the use of the narrative and evaluator's decision making model for creating a discourse between expert evaluators and novice evaluators as they discuss evaluation practice during evaluation training.

Key words: Program evaluation, Naturalistic Decision Making, Ill-structured problems, Situation analysis, Critical incidents, Critical Decision Making

Conducting program evaluations requires the use of many cognitive skills, such as problem solving and decision making. While there has been ample research on evaluation utilization (Patton, 2008), specifically to encourage stakeholders to make decisions based on evaluation results (Fleischer, & Christie, 2009) there is little empirical work on evaluators' practice, despite the strong call for "empirical studies on evaluation practice (Christi, 2009, p. 394)." Kundin (2008, 2010) has begun research on evaluator's decision making based on naturalistic decision making. This pilot study provides an example of the critical incidents that evaluators face when conducting evaluations and a model of the evaluator's decision making process. The narrative, generated from a retrospective interview, identified the evaluator's activities, decisions and the factors that influenced the evaluator's decisions.

Two assumptions were made during the pilot study. The first was that conducting an evaluation requires making one or more decisions. The second was that these decisions can be understood by researching the critical incidents faced by evaluators. Critical incidents are defined as significant events where one's expertise is challenged and where an event impacts the success or failure of individuals or systems in a given setting (Crandall, Klein, & Hoffman, 2006; Salas & Klein, 2001). An incident does not need to be a dramatic event. A critical incident can cause one to question an aspect of his or her beliefs, values, attitude, or behavior. In this study, critical incidents are defined as events that challenge an evaluator's performance and shape the outcome of an evaluation.

The purpose of this research is to answer the question: How do program evaluators make decisions when facing critical incidents? The question is broken down into two sub questions: What critical incidents did the evaluator face? What decisions did the evaluator make when faced with critical incidents?

Three conclusions have been determined from our findings. One conclusion is that evaluations are real-world ill-structured problems that are complex, contextually bound, and conducted under time constraints. The second conclusion is that problem structuring and problem solving are aspects of conducting an evaluation. The conclusion is that evaluators utilize at least one decision making approach: naturalistic decision making. This paper consists of the following sections: literature review, method, results, discussion, and conclusion.

Literature Review

This section provides a brief literature review on ill-structured problems and naturalistic decision making.

Ill-Structured Problems

A problem is defined as an unknown entity in some context where its initial state differs from its desired end state. A problem does not have a direct, obvious way to reach a needed solution (Jonassen, 1997). Problems can be viewed as well-structured problems or as ill-structured problems. A well-structured problem is defined as having "at least one problem space in which can be represented the initial problem state, the goal state, and all other states that may be reached or considered, in the course of attempting a solution to a problem" (Simon, 1973, p. 183) An ill-structured problem can be defined as a problem where the initial state's requirements are ambiguous, the specifications needed to define the solution are insufficient, and new constraints must be progressively defined during the solution generating process (Eastman, 1969; Jonassen, 1977; Simon, 1973). Often the goal is to move the ill-structured problem to a well-structured state. This is referred to as problem structuring (Smith, 1988). Problem solving involves moving a problem from its initial problem state to a desired end state.

Naturalistic Decision Making

Naturalistic decision making (NDM) research "asks how experienced people [or experts], working as individuals or groups in dynamic, uncertain, and often fast paced environments, identify and assess their situation, make decisions, and take actions whose consequences are meaningful to them and to the larger organization in which they operate" (Zsambok & Klein, 1997, p. 5). Experts or experienced proficient workers working in real-world situations must often perform in high-stress, time-pressured, high-stakes, interactive and complex, dynamic, or uncertain environments while working within an organization's goals and norms (Zsambok & Klein, 1997). As a result the expert may need to rapidly conduct a situation assessment and make real-time decisions on how to implement a workable course of action. An expert's naturalistic decision making is informed through situation assessments, contextual cues (Klein, 1997a) and results in satisficing (Simon, 1955), referred to as the production of a satisfactory solution rather than a solution that is optimal. This is referred to as the Recognition-Based approach to decision making. This is in contrast to the Rational Decision Making approach where decision makers gather all the available information, conduct exhaustive, concurrent analyses of the available options, and then choose an optimal solution (Hutton & Klein, 1999; Luce & Raiffa, 1957; Zannier, Chiasson & Maurer, 2007).

Method

Participant Description

This research was conducted during the month of July, 2010 after receiving Institutional Review Board (IRB) authorization. A participation request was sent to an evaluation organization based in the southeastern United States through the organization's listserv. An evaluator from the organization's health education evaluation special interest group responded to

the participation request. The interview took place in the southeastern United States, at the time and location most convenient for the interview subject. The participant was a female over 35 years of age who had completed a Master's degree in health education and promotion as well as a Ph.D. in the social sciences. The participant had at least ten years of program evaluation experience and had little to gain or lose from participating, at any level, in the study.

Procedure

In this case study one evaluator was interviewed. A retrospective semi-structured interview was utilized to elicit a critical incident event and to reveal the evaluator's decision making process. The interview lasted between about two hours. Content analysis was used to code words, phrases, sentences, and paragraphs. The resulting narrative answered the questions: What critical incidents did the evaluator face? What decisions did the evaluator make when facing critical incidents? Based on these findings, a narrative was generated to represent the critical incident events the evaluator faced. In addition, models were generated to illustrate the decision making process.

Data Collection

The interview was audio recorded and transcribed by the researcher. To stimulate discussions around evaluators' decision making, the Critical Decision Method (CDM) was used for "knowledge elicitation, data analysis, and knowledge representation" (Crandall, Klein, & Hoffman 2006, p. 9). The CDM was employed because it has been found conducive to examining and analyzing decisions (Salas & Klein, 2001). The CDM process began with a general question to initiate a conversation with the interview participant and to hear a description a critical incident. A semi-structured format (Patton, 2002) was utilized. During the interview, the evaluator was asked to describe a challenging program evaluation. Then timeline was created

with an emphasis on identifying instances of critical incidents and decision points. Probing questions were asked as a means of "deepening" (Crandall et al., 2006, p. p. 79) the interviewers understanding of the critical incident. An example of a probing question is "You mentioned that your department wanted the evaluation to track children's learning over time. How did you address that?" The interview participant was later asked to summarize how she made her decisions by answering open-ended questions such as "How did you make this decision?" More detail was sought by asking, "Anything else?" The post-decision summary (Ericsson & Simon, 1980) was used as a means of checking the validity and internal consistency of the data gathered through knowledge elicitation and knowledge representation process. After the interview was completed, the interviewee was provided an opportunity to ask questions about the study and was thanked for her participation. Two days after the interview, a narrative was generated based on stories that were put into a "(restoryied) into a framework that makes sense" (Creswell, 2007, p. 56). The narrative and timeline was reviewed by the research participant, the participant corrected any errors and the participant verified the final documents' accuracy.

Data analysis

Data analysis was conducted by analyzing the interview transcription, the narrative and the critical incidents/decision timeline. The interview transcripts were marked with a highlighter and labeled with codes found during data analysis (Krippendorff, 2004). Content analysis involved placing words, phrases, sentences, and/or paragraphs two categories: critical incident or decision. Then relevant passages of text were segmented. Coding small phrases assisted in revealing the participant's word choice in the transcription. Coding longer phrases, sentences, or paragraphs assisted in understanding context. To further understand the evaluator's decision making, relationships between the two codes were determined, as well as the relationship

between the critical incident and the decision. For example, the project manager's request to measure children's learning (a critical incident) was linked to the evaluator's decision not to measure children's learning over time (decision) due to lack of data (reason). The end evaluation was categorized as "satisficing", which is choosing the less than optimal option for a "realistic" option. Factors that influenced the program evaluator's decisions were also identified.

Results

During the CDM process, a timeline was co-created with the interviewee (see Figure A1.1). Prompts such as "Where in the timeline did this event occur?" were used for clarification purposes. The interviewee identified the critical incidents and the decision points during the construction of the timeline. The timeline became the foundation for the re-storyied narrative (see Appendix A1.2). As a result of the analysis, a table (see Table A1.1) was generated, linking each critical incident to a decision and their corresponding quotes. After studying the narrative, timeline, and critical incident-decision relationship table, factors that influenced the evaluator's decision making were labeled as "decision filters." A detailed evaluator's decision making diagram (see Figure A1.2) and a high-level overview of the evaluator's decision making process (see Figure A1.3) were generated.

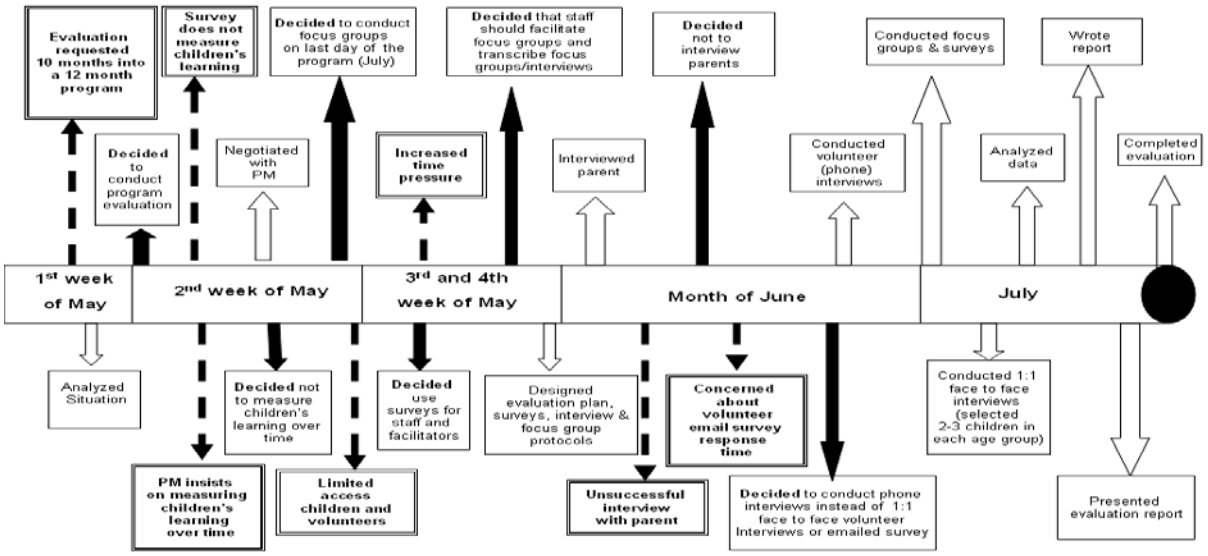

Figure A1.1 Evaluator's critical incidents, decisions, and timeline.

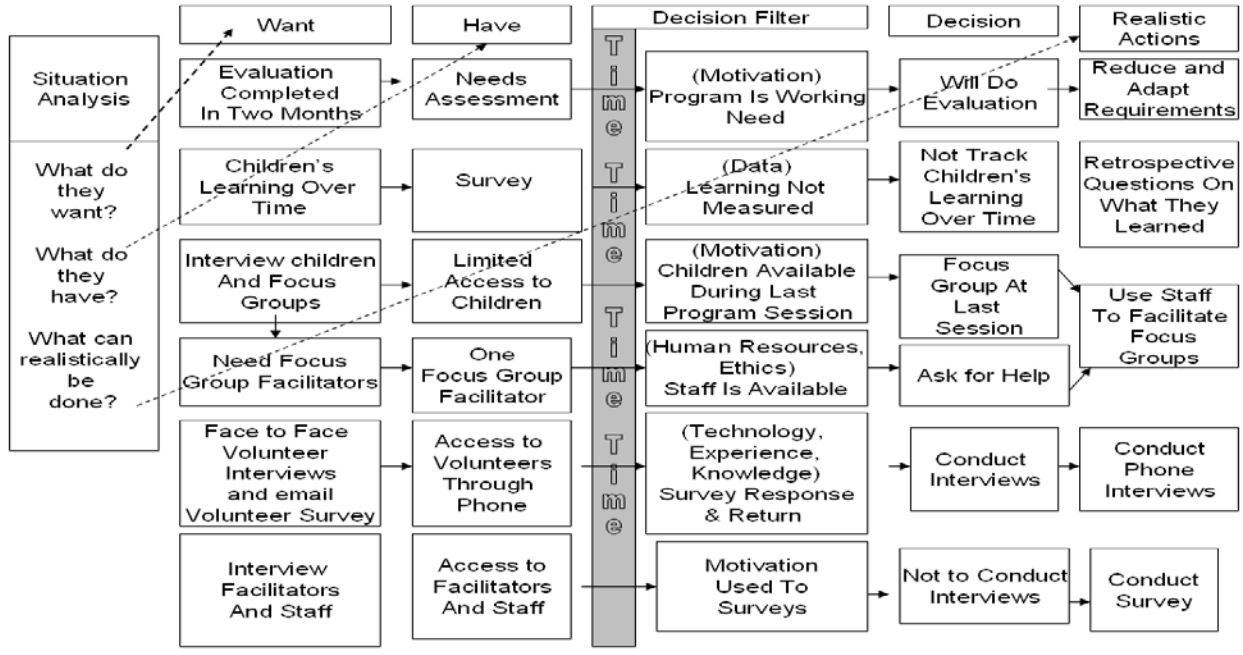

Figure A1.2. Detailed program evaluator's decision model.

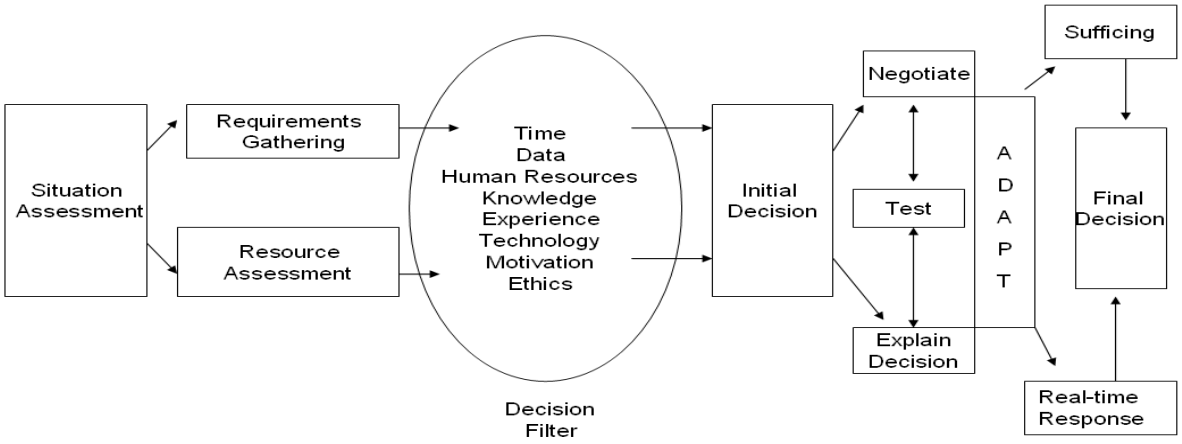

Figure A1.3. High-level overview of program evaluator's decision making process.

Table A1.1

Critical Incident-Decision Relationship with Corresponding Quotes

Critical Incident	Quote	Decision	Quote
Evaluation requested	You've asked me to evaluate the program ten months into the year-long program. The program is ending.	Decided to conduct program evaluation	I knew that the program needed to be evaluated
Ineffective Survey	The survey did not show if the children had learned anything.	Decided that measuring children's learning over time is not possible	We are not going to be able to track the kids over time.
Request to measure children's learning over time	"We want that."	Decided to do an end evaluation	"What we can do is an end evaluation that focuses on the different life skills.
Limited access to children and volunteers	I knew that it was hard enough to get the kids and volunteers to come to the sessions,	Made scheduling decisions	decided that it would be best to conduct the interviews during the final session of the year

Critical Incident	Quote	Decision	Quote
Increased time pressure	I couldn't be everywhere at the same time.	Decided that staff should facilitate focus groups and do transcriptions	The staff needs to facilitate two of the focus groups . . . [and] do the transcribing.
	I couldn't be everywhere at the same time.	Decided to use surveys for staff and facilitators	I decided that the facilitators and staff would complete an open-ended written survey.
	I know that the survey would have a low response rate . . .	Decided against email surveys for the volunteers	We decided to interview a few volunteers over the phone.
Unsuccessful parent interview	I called one parent and the interview was not productive.	Decided not to conduct interviews with parents	Conducting productive parent interviews would become a recommendation for the next evaluation

Contextual cues informed and factors influenced the decision making process. The evaluator's actions included conducting a situation assessment and the selection of solutions sufficient enough to get the job done: satisficing rather than selecting optimal solutions. The evaluator generated and implemented singular options (Zsambok & Klein, 1997) rather than generating a list of options and selecting one optimal solution. The evaluation environment, which was dynamic and complex, resulted in real-time reactions, negotiations with the project manager, explanations of her decisions, and adaptation of the evaluation plan to the changing evaluation context.

Evaluator Actions

Situation Analysis. The evaluator made her decisions through solution formulation, which included data collection, evaluating available resources, and constructing new information

(Brickell, Ferry & Harper, 2002; Xun & Land, 2004). The evaluator's situation analysis was conducted by asking questions throughout the process. The evaluator's goal was to generate "realistic" solutions. In this case study, the evaluator solicited problem clarification (Brickell, et al., 2002; Xun & Land, 2004). For example, the evaluator asked the project manager, "What is your evaluation plan?" The evaluator also asked herself, "What do you [they] already have? Let's see what's on the table." After finding out that an evaluation plan did not exist, the program evaluator decided "to help them figure out a *realistic* evaluation plan."

Satisficing. After the evaluator conducted a situation analysis, the evaluator utilized satisficing which as defined by Visser (2002, p. 3) involves "looking for good, acceptable, satisfactory solutions rather than one and only best, optimal solution from all possible" solutions This definition of satisficing comes from seminal research which focused on understanding ill-structured problems (Simon, 1995, p. 246; Simon, 1999). In this case study, the evaluator used cues as leverage decision points (Endsley & Garland, 2000; Klein, 1998). For example, the program director said, "We are already doing evaluation, so you can just capitalize off of that,'" to determine if the children had learned life skills." The evaluator determined that the surveys being used did not measure the children's learning. As a result, the evaluator to concluded that "[the program manager] didn't really understand it [the evaluation process] from the beginning." Due to time constraints and due to the lack of monitoring data, the evaluator decided that children's learning could not be measured. Instead, as a satisfactory solution, it was decided to use retrospective questions during the last program session to gather indications that children had learned life skills.

Singular Option. The evaluator did not generate a list of options but rather generated one option that she felt was "realistic." She did not spend a great deal of time on generating

options. This may have been due to the time constraints present throughout the evaluation process.

Real-time Reaction. The evaluation was conducted under significant time constraints. This resulted in real-time reactions such as deciding not to conduct parent interviews due to lack of time and expertise. The evaluator made other quick decisions such as the decision whether to conduct the evaluation.

Negotiation and Communication. In naturalistic decision making, the decision is not always conducted by one individual, but in collaboration with others. This occurred when the evaluator negotiated with the program manager. The evaluator stated, “We talked, back and forth, back and forth. I had to explain what could or couldn’t be done and why.” This indicates that the evaluator’s ability to negotiate and to communicate her definition of “realistic” was an essential skill.

Test, Explain, Negotiate, and Adapt. The evaluator decided to interview a parent whose child attended the program. According to the evaluator, the “interview was not productive.” The evaluator convinced the program manager that it would be “*impossible*” to interview the parents due to lack of time and expertise, since the “parents have [had] issues.” The evaluator explained the situation to the program manager. The program manager agreed that parents should not be interviewed. The evaluator and the stakeholder adapt to the situation by accepting the lack of parent interviews as a limitation of the study.

Decision Filters

Many factors influenced the evaluator's decision making. These factors were labeled as “Decision Filters.” The factors that influenced the evaluator’s decision making were: time

constraints, data, human resources, experience, knowledge, technology, and motivation. How these factors affected the evaluator's actions or decision making is further discussed.

Time constraints. Lack of time influenced the decision making process. This concern is highlighted this as a concern when the evaluator thought, “You’ve asked me to evaluate the program ten months into the year-long program. The program is ending.” The lack of time also influenced when the focus groups were scheduled, whether to conduct phone interviews instead of emailing surveys, and the request for staff support.

Data. The evaluator was very concerned about the lack of baseline data: the quality of the data being gathered as well as the tools being used to gather the data. For example, the evaluator stated that “there is [was] no baseline...nothing.” The evaluator also indicated it was necessary to “to read this *big* needs assessment’ *before* (interviewee’s emphasis) I [she] could begin to pull together an evaluation plan.” Concerns were also expressed about the organization’s children’s survey. It was stated, “The survey asked the children what they thought of the program, how did they like the program, and what did they think of the teacher. The children recorded their opinions by selecting a cartoon face that showed a frown, a straight mouth, or a smile.” The evaluator stated that the “survey didn’t show if the children had learned anything.”

Human Resources. The evaluator needed help from the program staff. The evaluator stated, “I couldn’t be everywhere at the same time. I let the director know that I needed help. . . “The staff needs to facilitate two of the focus groups . . .” and the staff needs to transcribe the field notes.

Experience. The evaluator indicated that the interview “with the parents didn’t work out” because “it would require more time and expertise than what was available.” The lack of experience in interviewing parents that “had issues” resulted in the decision not to conduct parent

interviews. Her experience as an evaluator guided her throughout the evaluation process to help her decide what can or cannot be done.

Knowledge. The evaluator used her background in health education and her expertise as an evaluator to help her design an evaluation plan, implement the plan, analyze the data, and report the findings. The evaluator relied on previous evaluation experience. This is implied by statements such as “I knew that an email survey would have a low response rate.”

Technology. The evaluator had concerns about utilizing email to send out the volunteer surveys. The evaluator was concerned that there may be a “low response rate and [that] the volunteers wouldn’t turn it [the survey] in on time.” Instead, the evaluator decided to interview a few volunteers over the phone. Phone interviews were chosen instead of using email to send out surveys.

Motivation. Motivation influenced on the evaluator's decision making. For example, when the evaluator decided to conduct the program evaluation she stated, “I decided to do it because I saw positive changes. I knew that the program needed to be evaluated to see what kind of impact it was having on the kids.” This statement indicates that because she saw positive results from the program, she was motivated to conduct the program evaluation.

Ethics. While conducting the evaluation, the evaluator checked for bias and conflict of interest. For example the evaluator stated, “I was worried about my potential bias since I was a facilitator for the program.” The evaluator also stated, “At first I was worried about a possible conflict of interest and that the children would try to give the staff positive answers just to please them. So I decided to use staff that didn’t normally work with the children. That was the best I could do and I listed it as a possible limitation in the evaluation report.”

Discussion

This study provided an opportunity to understand a program evaluator's naturalistic decision making and the critical incidents that they face. The evaluator's description of the critical incidents revealed the evaluator's actions and the factors that influenced the evaluator's decisions. The narrative revealed that program evaluators must solve ill-structured problems. In this case, the ill-structured problem began when the project manager requested an evaluation without an existing evaluation plan or data (initial state) without a clear definition of what the final evaluation should resemble (end state). Due to the context where the evaluation was conducted, the evaluation plan was continually modified. The influence factors of time, data, human resources, experience, knowledge, technology, motivation and ethics resulted in the need to be flexible, quick, resilient, adaptive, risk-taking and accurate. These are all characteristics identified by Cannon-Bowers and Bell (1997) as characteristics of effective decision makers. These characteristics seem to be necessary to conduct a real-world decision loaded evaluation. These characteristics may be necessary due to the critical incidents that evaluators face when conducting an evaluation.

Challenges and Benefits of the Critical Decision Method

Klein (2009) stated that "People become experts by the lessons they draw from their experiences, and by the sophistication of their mental models of how things work." Yet, as Salas and Klein stated, "memories cannot be assumed to be perfectly reliable" (2001, p. 99). This has caused NDM researchers to express apprehension about "the validity of verbal reports" (Wong, B. L., 2006, p. 3067-3075) since CDN utilizes a retrospective interview as a means of gathering information. Since people tend to recall the most recent incident this may prevent researchers from uncovering older aspects of a narrative.

The Critical Decision Method is a labor and time-intensive process (Salas & Klein, 2001, p. 99) for both the interviewer and the interviewee. This may be a deter experts from participating in interviews that utilize this approach. It is also recommended that two interviewers work as a team to conduct a CDM interview (Crandall et al., 2006; Salas & Klein, 2001). The cognitive load on from such an intensive interview may need to be offset by utilizing the assistance of another interviewer. This may not always be possible and therefore important information or cues may be missed. In addition, interviewers or interviewees can be overwhelmed by lengthy incidents. This may result in a compromised recall of the critical incidents that occur in an event (Wong, 2006).

CDM has been proven to be very effective in revealing cognitive elements from critical incidents and in revealing details regarding the decision making process. CDM has been used to build expert systems, to evaluate expert systems, and to determine training requirements. CDM provides opportunities for verification of information, probing for further clarification, modeling of decision making processes, and identifying critical incidents that occur in professional practice. The CDM has been used to evaluate decision making and to develop or improve training based on the researcher's findings (Freeman & Cohen, 1998; Grassi, 2000; Klein, 1997; Salas & Klein, 2001).

Conclusion

Previous studies have found that "researchers in naturalistic decision making, in particular, have designed training and decision aids based on discrepancies in the knowledge representations and cognitive strategies used by more and less experienced decision makers" (Crandall & Getchell-Reiter, 1993; Cohen & Freeman, 1997; Cohen, Freeman, & Thompson, 1997, 1998; Cohen & Thompson, 2001; Klein, 1997; Pliske, McCloskey, & Klein, 2001 as cited

by Lipshitz, & Cohen, 2005). There is limited research on the critical incidents that evaluators face. There has also been limited research on evaluators' decision making, and in particular their resulting naturalistic decision making as a response to the critical incidents (Kundin, 2008, 2010). This paper uses the Critical Decision Method to understand program evaluators' decision making process. Future studies on expert evaluators' naturalistic decision making and the narratives generated by the Critical Decision Method can be used to design training for novice program evaluators (Klein, 1997a, 1997b). The narratives can serve as the foundation for discourses that tests "practical experience against theoretical concepts and vice versa" (Christie & Rose, 2003) which can result in learning about evaluation practice.

REFERENCES

Brickell, G., Ferry, B., & Harper, B. (2002, December). *Designing informal reasoning skills in ill-structured problems: A case study into problem solving strategies*. Paper presented at ASCILITE Conference, Auckland, New Zealand.

Cannon-Bowers, J. A. & Bell, H. H. (1997). Training decision makers for complex environments: Implications for the naturalistic decision making perspective. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 99–110). Mahwah, NJ; Erlbaum.

Christie, C. A. (2009). Analyzing the practice of evaluation. In J. Fitzpatrick, C. A. Christie, & M. M. Mark (Eds.), *Evaluation in action: Interviews with expert evaluators* (pp. 393–435). Thousand Oaks, CA: Sage.

Christie, C.A & Rose, M. (2003). Learning about evaluation through dialogue: Lessons from an informal discussion group. *American Journal of Evaluation, 24*(2), 235–243.

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: A technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science, 16*(1), 41–51.

Eastman, C. M. (1969) Cognitive processes and ill-defined problems: a case study from design. In: Walker DE, Norton LM (Eds.) *Proceedings of the international joint conference on artificial intelligence (IJCAI '69),* Washington, DC, May 1969, pp 675–699.

Endsley, M. R. & Garland, D. J. (2000). *Situation awareness analysis and measurement.* Mahwah, N.J: Lawrence Erlbaum Associates.

Fitzpatrick, J. L., Saunders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, NJ: Pearson.

Fleischer, D. & Christie, C.A. (2009). Evaluation use: A survey of U.S. American Evaluation Association members. *American Journal of Evaluation, 30*(2), 158–175.

Freeman, J. T., & Cohen, M. S. (1998). A Critical Decision Analysis of Aspects of Naval Anti-Air Warfare. *Cognitive Technologies, Inc., Technical Report 98–2*.

Grassi, C. R. (2000, September). *A task analysis of pier side ship-handling for virtual environment ship-handling simulator scenario development* (Master's thesis). Naval Postgraduate School. Monterey, CA. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA384256&Location=U2&doc=GetTRDoc.pdf

Hutton, R. J. & Klein, G. (1999). Expert decision making. *Systems Engineering, 2*(1), 32–45.

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology: Research and Development, 45*(1), 65–94.

Klein, G. A. (1997a). Developing expertise in decision making. *Thinking and Reasoning, 34*, 337–352.

Klein, G.A. (1997b). The recognition-primed decision (RPD) model: Looking back, looking forward, In C. E. Zsambok and G. Klein, *Naturalistic decision making* (pp. 285–292). New York, NY: Routledge.

Klein. G. A. (1998). *Sources of Power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. A. (2009). *Streetlights and shadows: Searching for the keys of adaptive decision making*. Cambridge, MA: MIT Press.

Krippendorff, K. (2004). *Content Analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Kundin, D. M. (2008). *Everyday approaches to evaluation: A study of how evaluators make practice decisions.* (Doctoral dissertation). Retrieved from ProQuest. AAT 3299404

Kundin, D. M. (2010, June). A conceptual framework for how evaluators make everyday decisions. *American Journal of Evaluation, 31*(3), 347–362

Luce, R. D, & Raiffa, H. (1957). *Games and decisions*. New York, NY: Wiley.

Patton, M. Q. (2008). *Utilization focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Salas, E. & Klein, G. A. (2001). *Linking expertise and naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*, 99–118.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence, 4*(3–4), 181–201.

Smith, G. F. (1988). Towards a heuristic theory of problem structuring. *Management Science, 34*(12), 1489–1506.

Visser, W. (2000, May). A tribute to Simon and some - too late – questions by a cognitive ergonomist. INRA Report 4462. Rocquencourt, France: INRIA

Wong, B. L. (2006). The critical decision method. In W. Karwoski (Ed.), International *Encyclopedia of Ergonomics and Human Factors* (Vol. 9) (pp. 3067–3075). Boca Raton, FL: CRC Press.

Xun, G., & Land, S. L. (2004). A conceptual framework for scaffolding ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development, 52*(2), 5–22.

Zannier, C., Chiasson, M., & Maurer, F. (2007). A model of design decision making based on empirical results of interviews with software designers. *Information and Software Technology, 49*, 637–653.

Zsambok, C. E. (1997). Naturalistic decision making: Where are we now? In C. E. Zsambok & Klein, G. (Eds.), *Naturalistic Decision Making* (pp. 3–15). New York, NY: Routledge.

Zsambok, C. E., & Klein, G. (1997). *Naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pilot Study Narrative

I evaluated a program that teaches life skills to children who are affected or infected by HIV/AIDS. The program meets with children once a month for two hours to teach life skills such as money management, communication skills, and how to handle their emotions. About ten children, a facilitator, and an adult volunteer are present at each of the three sessions. The sessions are divided into three age groups: 6-9, 10-13, and 14-18 years old.

One day, the program director approached me and said, "Oh, [I heard] you're an evaluator as well. Can you evaluate the project?" He requested this ten months into a twelve-month program. I knew that, like many programs, they had not included an evaluation plan when they developed the program.

My next thought was, "What do you already have? Let's see what's on the table." So I asked, "What is your evaluation plan?" And she said, "We don't have one." So I thought to myself, "Okay, you want me to evaluate this, and there is no baseline; there is nothing." I asked them, "What do you want to see out of this evaluation?" They said that one of the things that they want to do is to track each individual child's progress from the beginning to the end of the program. I thought to myself, "You've asked me to evaluate the program ten months into the year-long program. The program is ending." I decided to do it because I saw positive changes in the children. I knew that the program needed to be evaluated to see what kind of impact it was having on the kids. I knew that the program's funders needed the evaluation to determine if they would continue to fund the program

The program manager didn't really understand it from the beginning. The program director said, "We are already doing evaluation, so you can just capitalize off of that." The program director thought evaluations were being conducted when the children filled out a survey

at the end of each session. The survey asked the children what they thought of the program, how did they like the program, and what did they think of the teacher. Children reported their opinions by selecting a cartoon face that showed a frown, a straight mouth, or a smile. The survey did not show if the children had learned anything.

I had to find a way to tell them that the survey wasn't measuring learning over time. So, I said, "I can do something, but I can't do everything that you want. We are not going to be able to track the kids over time. You haven't collected any data over time, so you are not going to be able to do that." The director said, "We want that." Some negotiation had to be done to help them figure out a *realistic* evaluation plan.

The program director wanted me to interview all the kids from each age group. She also wanted me to interview parents, volunteers, facilitators, and staff. We talked, back and forth, back and forth. I had to explain what could or could not be done and why. That process took a lot of time, which meant that I no longer had the *entire* two months to conduct the evaluation. I kept thinking, "Okay, this is what they want. This is what they have. This is what can *realistically* be done in a month and a half."

The director wanted the evaluation to determine if the program met the needs of the children as determined by the needs assessment. I thought to myself, "Now, I have to read this *big* needs assessment." This had to be done *before* I could begin to pull together an evaluation plan.

I reviewed the needs assessment, the lessons taught through the program, and the program's objectives. I developed some surveys and baseline questions. The questions were retrospective. Although I knew that it would be difficult for the children to remember what had happened in the past, it was the best we could do. I asked the program manager, "What do you

want to know from the parents, the staff, the volunteers, and the facilitators?" I developed a couple of interview questions based on the director's responses.

Meanwhile, I was worried about my potential bias since I was a facilitator for the program. I think the evaluation worked out because I *knew* the program, which helped me to set realistic goals and to see what could or could not be done. That meant uncovering the *good, the bad, and the ugly* and providing recommendations so that the program could learn from the evaluation. Being familiar with the program also helped me to plan the interview schedule. I knew that it was hard to get the kids and volunteers to come to the sessions, and this insider information helped me to decide to conduct the interviews during the final session of the year.

The program director still wanted to track kids over time, but I told her that we would not be able to do that. I said, "What we can do is an end evaluation that focuses on the different life skills. We can interview a few of the youths."

I couldn't be everywhere at the same time. I let the director know that I needed help. I said, "The staff needs to facilitate two of the focus groups, and I will facilitate one focus group." Of course, I couldn't facilitate the focus group for which I was already a program facilitator. It was also decided that staff would need to do the transcribing. We just didn't have *enough* time and enough help. At first I was worried about a possible conflict of interest and that the children would try to give the staff positive answers just to please them. So I decided to use staff that didn't normally work with the children. That was the best I could do and I listed it as a possible limitation in the evaluation report.

Regarding the volunteers, the program manager decided that she wanted a focus group and an email survey sent to the volunteers. I knew that an email survey would have a low

response rate and that the volunteers wouldn't turn them in on time. We decided to interview a few volunteers over the phone.

I decided that the facilitators and staff would complete an open-ended written survey. I knew that they would turn in the surveys on time. We just didn't have time to do any one-on-one interviews.

Sadly, the interviews with the parents didn't work out. I called one parent and the interview was not productive. A lot of the parents have issues. I wasn't getting *good data.* It would require more time and expertise than what was available. I had to be *realistic*. The program director agreed that it would be *impossible* to conduct interviews with the parents, though it would be a big limitation of the evaluation. Conducting productive parent interviews would become a recommendation for the next evaluation.

Eventually, the evaluation started to flesh out very nicely. We got some good information from the kids and from the volunteers. The program directors found out that volunteer training wasn't included in the curriculum. They learned about the challenges the staff encountered and how the needs of the staff could be met. The program facilitators gave good feedback and confirmed many of the findings from the other groups.

The program director also learned that scheduling prevented some of the kids from participating. The sessions were held on Saturdays from 2-5 p.m. Program participants felt this was an odd time, as it was in the middle of the day and they couldn't do very much after that. To make it worse, the facilitator had to be there an hour before each session. Participants also felt that the sessions were too long.

The evaluation results definitely informed the next cycle. I have been with them for three years, and I have seen the changes as well. After I put everything together and presented the

report to them, they were amazed at the amount of good information. It was only then that the program director said, "Now, I see what you mean when you said that we really *do* need to track the kids from the beginning." It was an "ah-ha moment."

The program director and the program funders needed to be *satisfied* with the evaluation we were able to do, given the circumstances and limitations we had. After I put everything together and presented the report to the program director & funders, they were amazed at the amount of good information collected. It was only then that the program director said, "Now I see what you mean when you said that we really do need to track the kids from the beginning." It made them see, "Oh wow, we needed to start this evaluation at the beginning." The decided that "For the next cycle, let's start with our evaluation plan while we are developing this program."

These evaluation results definitely informed the next program cycle. I have been with this program for three years now, and I have seen the changes. Now, they think about "What data gathering instruments should we use?" The participant survey forms are different now. Instead of questions that ask the participant to choose smiley faces, the questions ask for concrete information like "What things did you learn?"

Figure A2.1. Narrative based on Pilot Study: A Program Evaluator's Decision Making When Facing Critical Incidents.

Pilot Study Interview Protocol and Questions

The University of Georgia

Interview Protocol for a Study on

Factors that Influence Expert HIV/AIDS Health Education Program Evaluators Decision Making during Critical Incidents: Using the Critical Decision Method

To facilitate note-taking, I would like to audio record our conversations today. For your information, only researchers on the project will be privy to the recordings, which will be destroyed after transcription is completed.

Introduction: The purpose of this study is to examine the factors that influence expert evaluators' decision making and the nature of program evaluation tasks. Specifically, this study will investigate how expert evaluators of HIV/AIDS Health Education programs arrive at decisions when they encounter non-routine, challenging events while evaluating an HIV/AIDS health education program that requires them to use their expertise to make decision that will affect the outcome of their program evaluation. This study aims to understand your reasoning as you solve problems that may have occurred when conducting an evaluation. Your recall of critical decision making incidents and your insights into expert evaluators' decision making will inform this study.

Sweep 1: Explain research study (in introduction) and consent forms. Obtain volunteer participant's consent.

Possible interview questions and review of participant's resume

Background information

1. How many years of experience do you have in evaluating HIV/AIDS health education programs?
2. How many of these programs were situated in Sub-Saharan Africa?
3. How many of these programs were funded by international aid agencies?

Sweep 2- obtaining an incident, ask for an overview:

1. Can you think of a time when your skills as a HIV/AIDS program evaluator were really challenged?
2. Can you share with me a time when your skills as an evaluator of HIV/AIDS programs really made a difference?
3. Can you share with me an instance when a HIV/AIDS program evaluation would have gone differently if you weren't there?
4. Can you tell me about the last time your decision making skills as an evaluator were challenged?

Sweep 3 – Timeline Co-Construction, Verification and Decision Point Identification

1. Where in the time line should I put this?
2. Do I have this right?
3. Where in the timeline do you think this critical incident occurred?
4. When do you feel that this decision had to be made?
5. Where do you think that this decision resolved the critical incident?

Sweep 4- Deepening

1. What was the situation that let you know what was going to happen?
2. What was it about the situation that let you know what to do?
3. What led up to this decision?
4. What were your concerns at this point?
5. How would you summarize the situation at this point?
6. What information did you use to make this decision? How did you get this information?
7. What evaluation knowledge was necessary or helpful in this situation?
8. What were your specific evaluation goals at this time?
9. What were you hoping as an evaluator to accomplish at this point?

Sweep 5 – "What if" queries

1. What alternatives did you consider?
2. What would another evaluator possibly do differently?
3. What other actions could you have taken?
4. How would you have approached this decision earlier in your evaluation career?
5. How would this incident turned out differently if someone without your level of expertise had not been there?
6. If you were to encounter this type of incident again, how would you approach it?

Sweep 6 – Clarify and Validate Narrative - Conclude Interview

1. (While summarizing the critical incident story, timeline and reviewing the key factors) What other observations would you like to add regarding your decision making process?
2. What other observations would you like to add regarding the critical incident(s)?
3. Did I miss understand this? Did I forget something?
4. Is there something you would like to add?

Figure A3.1. Pilot study interview protocol and questions.

Pilot Study Consent Form

CONSENT FORM

I agree to take part in a research study titled* "Factors that Influence Expert HIV/AIDS Health Education Program Evaluators' Naturalistic Decision Making During Critical Incidents: Applying the Critical Decision Method"** ***which is being conducted by **Ms. Anita F. Zgambo, Educational Psychology and Instructional Technology, University of Georgia, 404-667-9631, *under the direction of* Major Advisor, Dr. Ikseon Choi, Educational Psychology and Instructional Technology, University of Georgia (706-583-0794).** ***My participation is voluntary; I can refuse to participate or stop taking part at any time without giving any reason, and without penalty. I can ask to have information related to me returned to me, removed from the research records, or destroyed.***

Title of Research Study: Factors that Influence Expert HIV/AIDS Health Education Program Evaluators' Naturalistic Decision Making During Critical Incidents: Applying the Critical Decision Method

1) Reason or Purpose:

This study will utilize cognitive task analysis to identify the factors that influence expert evaluator's decision making when they encounter non-routine situations that require expert decisions or judgment while evaluating Sub-Saharan HIV/AIDS health education programs funded by international development agencies. The aim of this research is to study the factors that influence expert evaluators' decision making and the nature of evaluation tasks. The research will result in the identification of factors that influence expert evaluators' decision making based on semi-structured interview generated narratives and the future development of

curriculum that will be utilized to teach evaluation concepts through an e-learning portal. The interviews will also reveal the challenging aspects of conducting evaluation tasks. Your recall of critical decision making incidents and your insights into expert evaluators' decision making will inform this study.

2) Benefits:

To take part in this research, the participants will have a chance to reflect on their experiences, decision making and the nature of evaluation tasks. The questions that asked may have provide evaluators the opportunity to reflect on their decision making process and skills. This may benefit their future program evaluation skills. The knowledge gained from this study will be utilized in the future to generate a educational case library of realistic program evaluation experiences. The case library will contribute the program evaluation curriculum development that will utilize a e-learning format.

3) Procedures:

To participate in the current study need to have conducted program evaluations of HIV/AIDS health education programs. The process involves up to three one-hour interviews. You will be provided a copy of the resulting narrative based on your recalled account in order to verify its accuracy. The interviews will be audio-recorded and usually will last for one hour. All personal identifiers will be removed from the transcripts and the audio recordings will be destroyed after the transcription has been completed. It is anticipated that in the future realistic case developed from this study, which will be used to train future program evaluators.

4) **Discomforts or Stress:**

No discomforts or stresses are expected.

*5) Risks***:**

No risks are expected.

6) Data Collection & Storage:

The only people who will know that I am a research subject are members of the research team. No individually-identifiable information about me, or provided by me during the research, will be shared with others, except if necessary to protect my rights or welfare (for example, if I am injured and need emergency care); or if required by law.

Skype may be used as a means of Internet communication. Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However once the materials are received by the researcher, standard confidentiality procedures will be employed.

I have the right to review/edit the tapes and only the researchers will have access to the tapes. Audio recordings will be destroyed after transcription has been completed. While responses may be linked back to an individual participant by the researchers, all data will be labeled with a key code. The codes will be secured in a locked container in a locked room. Coded data will be maintained in a different location. Computer files will be used but the keys to coded data will be kept in an encrypted and/or password protected file. The coded data file will be maintained on a separate computer.

7) **Contact Information:**

The researcher will answer any further questions about the research, now or during the course of the project, and can be reached by telephone at: 404-667-9631.

8) Consent Statement:

My signature below indicates that the researchers have answered all of my questions to my satisfaction and that I consent to volunteer for this study. I have been given a copy of this form.

_____ _____ _____

Name of Researcher Signature Date

Telephone: 404-667-9631

Email: azgambo@gmail.com or azgambo@uga.edu

_____ _____ _____

Name of Participant Signature Date

Please sign both copies, keep one and return one to the researcher

Additional questions or problems regarding your rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 629 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu

Figure A4.1. Pilot study consent form.

APPENDIX B

IRB FORM

Institutional Review Board
Human Subjects Office
612 Boyd GSRC
Athens, GA 30602-7411

The University of Georgia

Phone: 706-542-3199
Fax: 706-542-3360
Email: irb@uga.edu

IRB CONTINUING REVIEW/AMENDMENT FORM

Principal Investigator (PI): Dr. Ikseon Choi	
Co-Principal Investigator (Required, if co-PI is a student): Anita Frances Zgambo	
Project #: 2010-10839-0	Title of Study: Factors that Influence Evaluators' Naturalistic Decision Making When Evaluating An HIV/AIDS Health Education Program Implemented in Sub-Saharan Africa and Funded by International AID Agencies: Utilizing the Critical Decision Method - changed to - Modeling Evaluator's Naturalistic Decision Making when Evaluating HIV/AIDS Health Education Programs Implemented in Sub-Saharan Africa and Funded by International Aid Agencies

	PLEASE ANSWER ALL QUESTIONS (Use the text boxes for explanation/additional information or attach a separate cover letter.)	YES	NO
1	Have you started data collection for this research project?	☒	☐
2	How many total participants have been accrued since the beginning of the research project? (*Note: This corresponds to the number of individuals who gave consent; this number should include withdrawals but actual number of withdrawals is reported in #7 below.*)	1	
3	Do you plan to continue to recruit participants for this research project? (If you answered **YES**, please skip to Question #6.)	☒	☐
4	If you answered **NO** to question #3, do you plan to continue to collect data with previously recruited participants?	☐	☐
5	If you answered **NO** to questions #3 and #4 above, do you plan to continue to analyze previously collected data that is individually-identifiable?	☐	☐
6	Have there been any complaints about the research since the protocol was approved by the IRB? If **YES**, please provide complete information on the complaints made.	☐	☒
7	Have any participants withdrawn, dropped out, or were lost to follow-up from participation since the protocol was last approved by the IRB? If **YES**, please indicate the number and provide detailed information/reason(s).	☐	☒
8	Have there been any adverse events or unanticipated problems involving risks to the participants or others since the protocol was last approved by the IRB? If **YES**, please contact the IRB office immediately to request an adverse event/incident report form.	☐	☒
9	Have there been any changes to the study population? If **YES**, please explain changes.	☐	☒
10	Have the procedures changed in any way since the protocol was last approved by the IRB? If **YES**, please explain. **Procedure has changed from 3 one-hour interviews to 2 one-hour interviews. E-learning portal will not longer be part of the study but may be developed in the future to train program evaluators**	☐	☐
11	Have any materials or instruments changed in any way since the protocol was last approved by the IRB? If **YES**, please explain. **The consent form, recruitment communication and interview protocal has been changed to reflect the changed research title, the reduced number of interviews (from 3 to two one-hour interviews) and the development of a e-learning portal for future development.**	☒	☐
12	Have changes in the scientific literature, or interim experience with this or related studies, changed your assessment of potential risks or benefits to study participants? If **YES**, please explain and attach any relevant literature.	☐	☒
13	Have the consent documents changed in any way since the protocol was last approved by the IRB? If **YES**, please explain and attach copy of the revised document(s). **Documents attached**	☒	☐
14	A clean copy of the current version of the consent document(s) must be submitted with the request to continue **if you plan to recruit new participants, or if a revised consent document is necessary as a result of an amendment**. Have you attached a clean copy of your current consent document(s)?	☒	☐

15	Have there been any changes to the members of the research team (e.g., change in PI; addition/deletion of co-investigators)? If **YES**, please describe personnel change(s). Note: All new personnel must complete the CITI training.	☐	☒
Principal Investigator's Signature: **For electronic submission, a check in this box is acceptable as a signature:** ☒		**Date:** **01/31/2011**	

Important: If research activities involving human participants will continue five years after the original IRB approval, please submit a new IRB Application for initial review. **Exceptions**: If the research is permanently closed to the enrollment of new subjects, all participants have completed all research-related interventions, **and** the research will remain active only for long-term follow-up of subjects; or if the remaining research activities are limited to analysis of individually-identifiable private information.

APPENDIX C

CONSENT FORM

I agree to take part in a research study titled "Modeling Evaluators' Naturalistic Decision Making When Evaluating an HIV/AIDS Health Education Program Implemented in Sub-Saharan Africa and Funded by International Aid Agencies" *which is being conducted by* Ms. Anita F. Zgambo, Educational Psychology and Instructional Technology, University of Georgia, 404-667-9631, *under the direction of* Major Advisor, Dr. Ikseon Choi, Educational Psychology and Instructional Technology, University of Georgia (706-583-0794). *My participation is voluntary; I can refuse to participate or stop taking part at any time without giving any reason, and without penalty or loss of benefits to which I am otherwise entitled. I can ask to have information related to me returned to me, removed from the research records, or destroyed.*

Title of Research Study

Factors that Influence Expert HIV/AIDS Health Education Program Evaluators' Naturalistic Decision Making During Critical Incidents: Applying the Critical Decision Method

1) Reason or Purpose: This study will utilize cognitive task analysis to identify the factors that influence expert evaluator's decision making when they encounter non-routine situations that require expert decisions or judgment while evaluating sub-Saharan HIV/AIDS health education programs funded by international development agencies. The aim of this research is to study the factors that influence expert evaluators' decision making and the nature of evaluation tasks. The research will result in the identification of factors that influence expert evaluators' decision making based on semi-structured interview generated narratives and development of curriculum

that will be utilized to teach evaluation concepts through an e-learning portal. The interviews will also reveal the challenging aspects of conducting evaluation tasks. Your recall of critical decision making incidents and your insights into expert evaluators' decision making will inform this study.

2) Benefits: To take part in this research, the participants will have a chance to reflect on their experiences, decision making and the nature of evaluation tasks. The questions that asked may have provide evaluators the opportunity to reflect on their decision making process and skills. This may benefit their furture program evaluation skills. In the future and after the completion of this dissertation study, the knowledge gained from this study will generate a educational case library of realistic program evaluation experiences. The resulting case library will contribute to curriculum development that will utilize a e-learning format that will be used to provide professional development/educational training for current or aspiring evaluators.

3) Procedures: To participate in the current study need to have conducted program evaluations of HIV/AIDS health education programs. The process involves two one-hour interviews. You will be provided a copy of the resulting narrative based on your recalled account in order to verify its accuracy. The interviews will be audio-recorded and usually will last for one hour. All personal identifiers will be removed from the transcripts and the audio recordings will be destroyed after the transcription has been completed. Any codes that can link you to the data will be destroyed after the transcription has been completed. A realistic case will result from this study, which will be used to train future program evaluators. The realistic case will be available for you to review.

4) **Discomforts or Stress:** No discomforts or stresses are expected.

5) Risks: No risks are expected.

6) Data Collection & Storage: The only people who will know that I am a research subject are members of the research team. No individually-identifiable information about me, or provided by me during the research, will be shared with others.

Skype may be used as a means of Internet communication. Internet communications are insecure and there is a limit to the confidentiality that can be guaranteed due to the technology itself. However once the materials are received by the researcher, standard confidentiality procedures will be employed. If a second interview cannot be secured, through telephone or Skype, the narrative resulting from the first interview will be emailed for verification with the questions for the second interview embedded into the document.

I have the right to review/edit the tapes and only the researchers will have access to the tapes. Audio recordings will be destroyed after transcription has been completed. While responses may be linked back to an individual participant by the researchers, all data will be labeled with a key code. Any codes that can link you to the data will be destroyed after the interview transcription has been completed. Meanwhile, the codes will be secured in a locked container in a locked room. Coded data will be maintained in a different location. Computer files will be used but the keys to coded data will be kept in an encrypted and/or password protected file. The coded data file will be maintained on a separate computer.

7) Contact Information: The researcher will answer any further questions about the research, now or during the course of the project, and can be reached by telephone at: 404-667-9631.

8) Consent Statement: My signature below indicates that the researchers have answered all of my questions to my satisfaction and that I consent to volunteer for this study. I have been given a copy of this form.

_____ _____ _____

Anita F. Zgambo Signature Date

Telephone: 404-667-9631 or 202-470-5779

Email: azgambo@gmail.com or azgambo@uga.edu

_____ _____ _____

Name of Participant Signature Date

Please sign both copies, keep one and return one to the researcher

Additional questions or problems regarding your rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 629 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu

APPENDIX D

DECISION MAKING TIMELINES AND MODELS

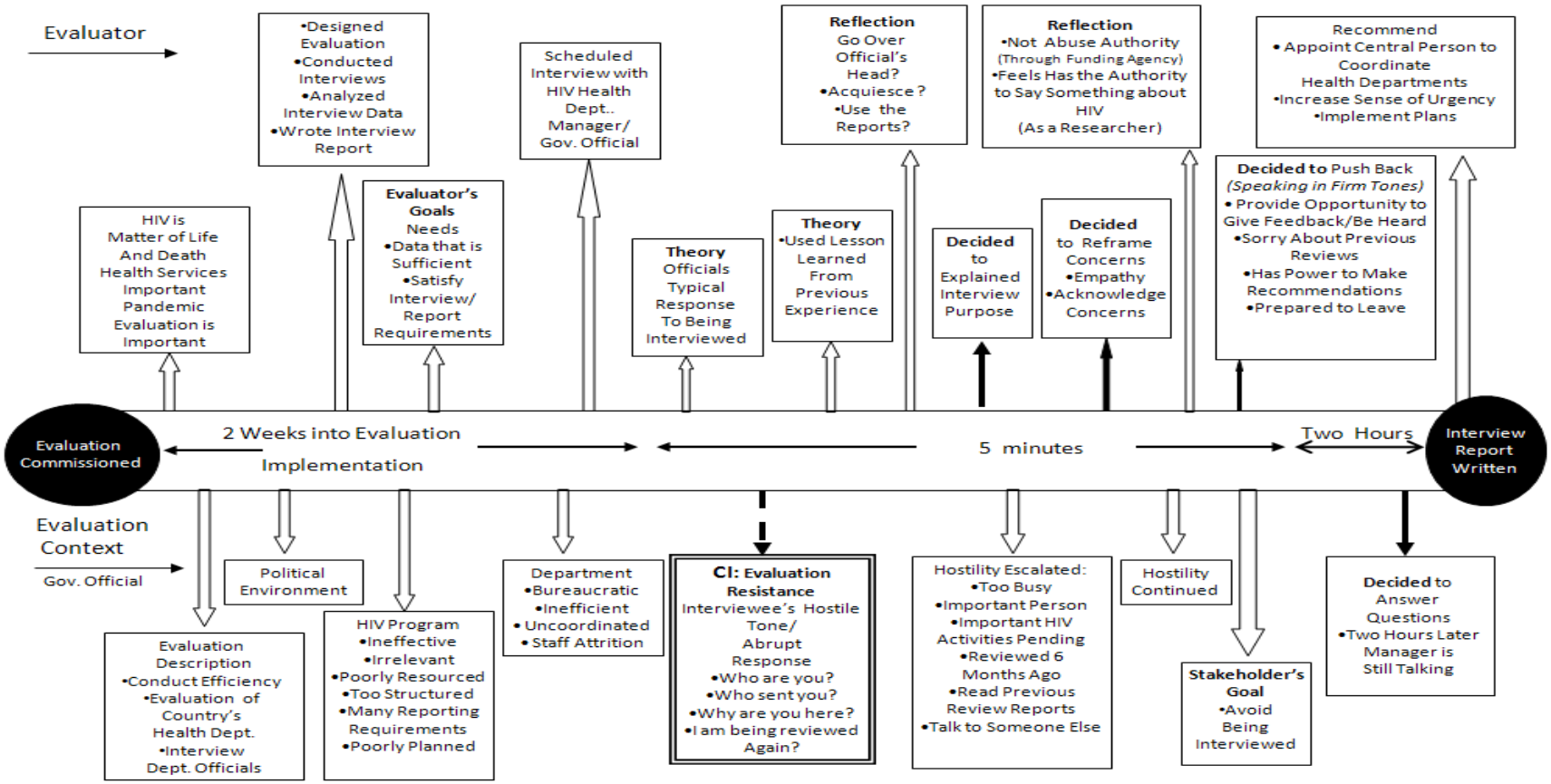

Figure D1.1. Case 1 Decision Making Timeline - Formative evaluation of an HIV health department and an interview of an evaluation department official.

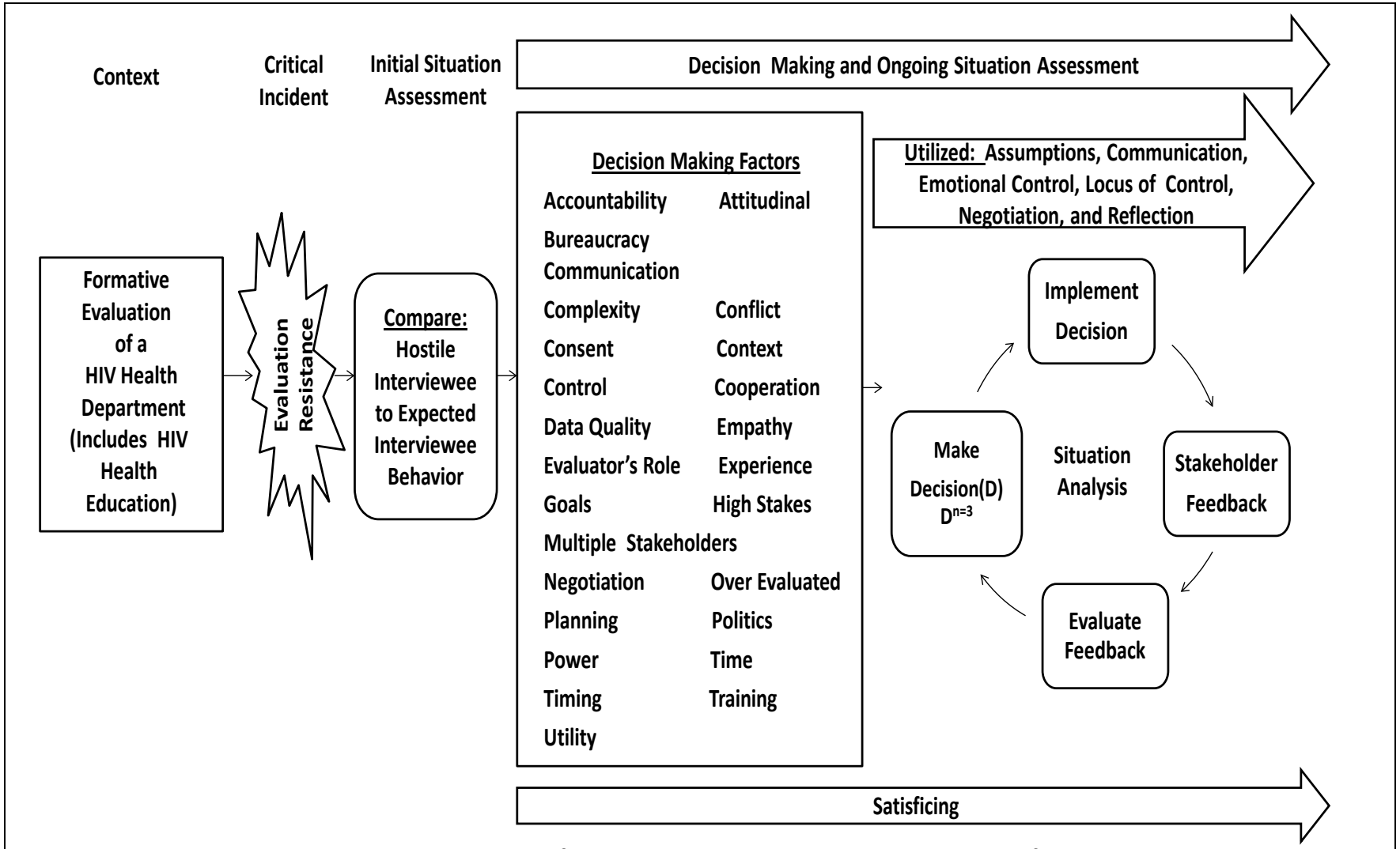

Figure D1.2. Case 1 Model - Formative evaluation of an HIV health department and an interview of an evaluation department official.

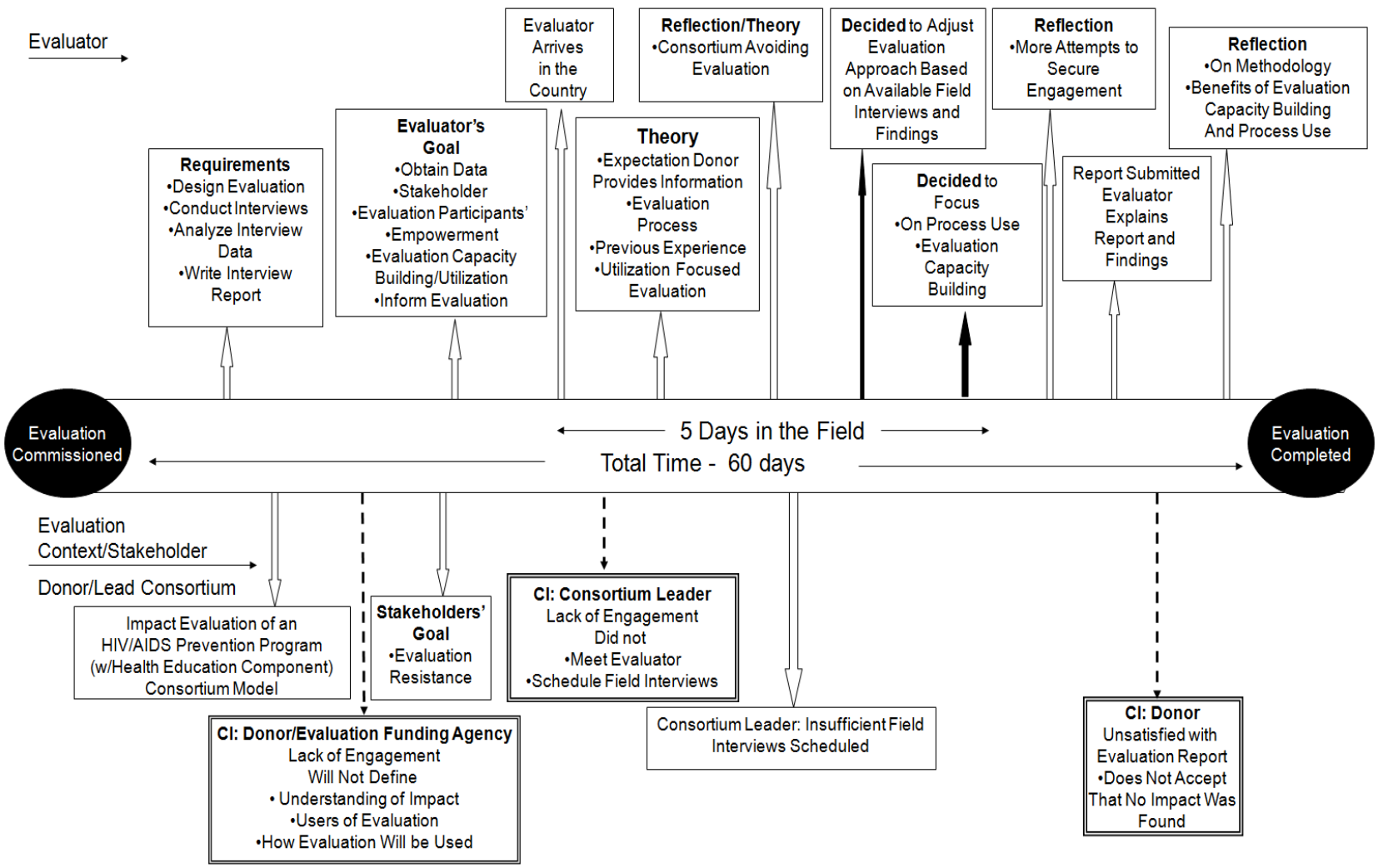

Figure D1.3. Case 2 Decision Making Timeline - Summative impact evaluation of an HIV/AIDS health education program's consortium model.

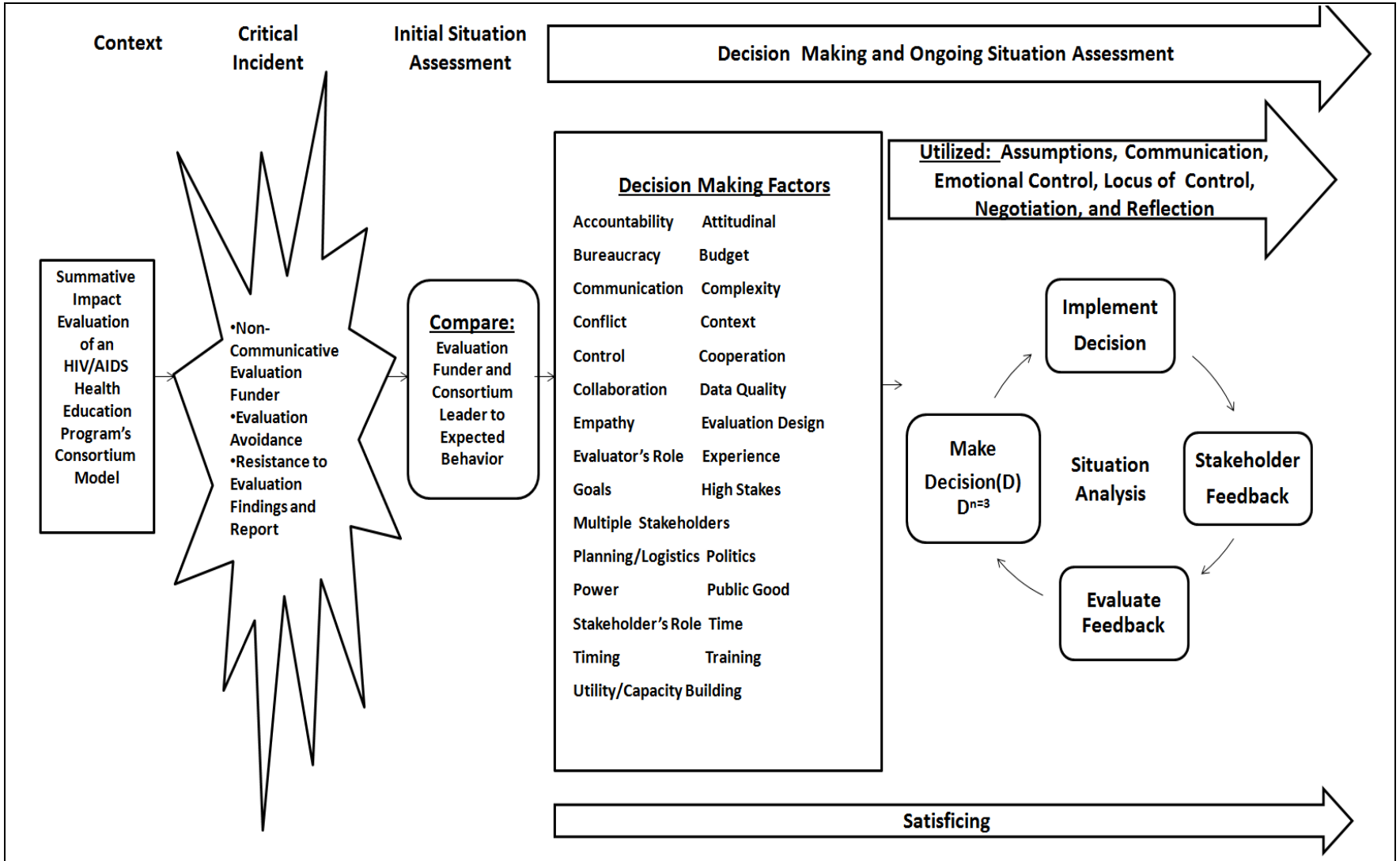

Figure D1.4. Case 2 Model - Summative impact evaluation of an HIV/AIDS health education program's consortium model.

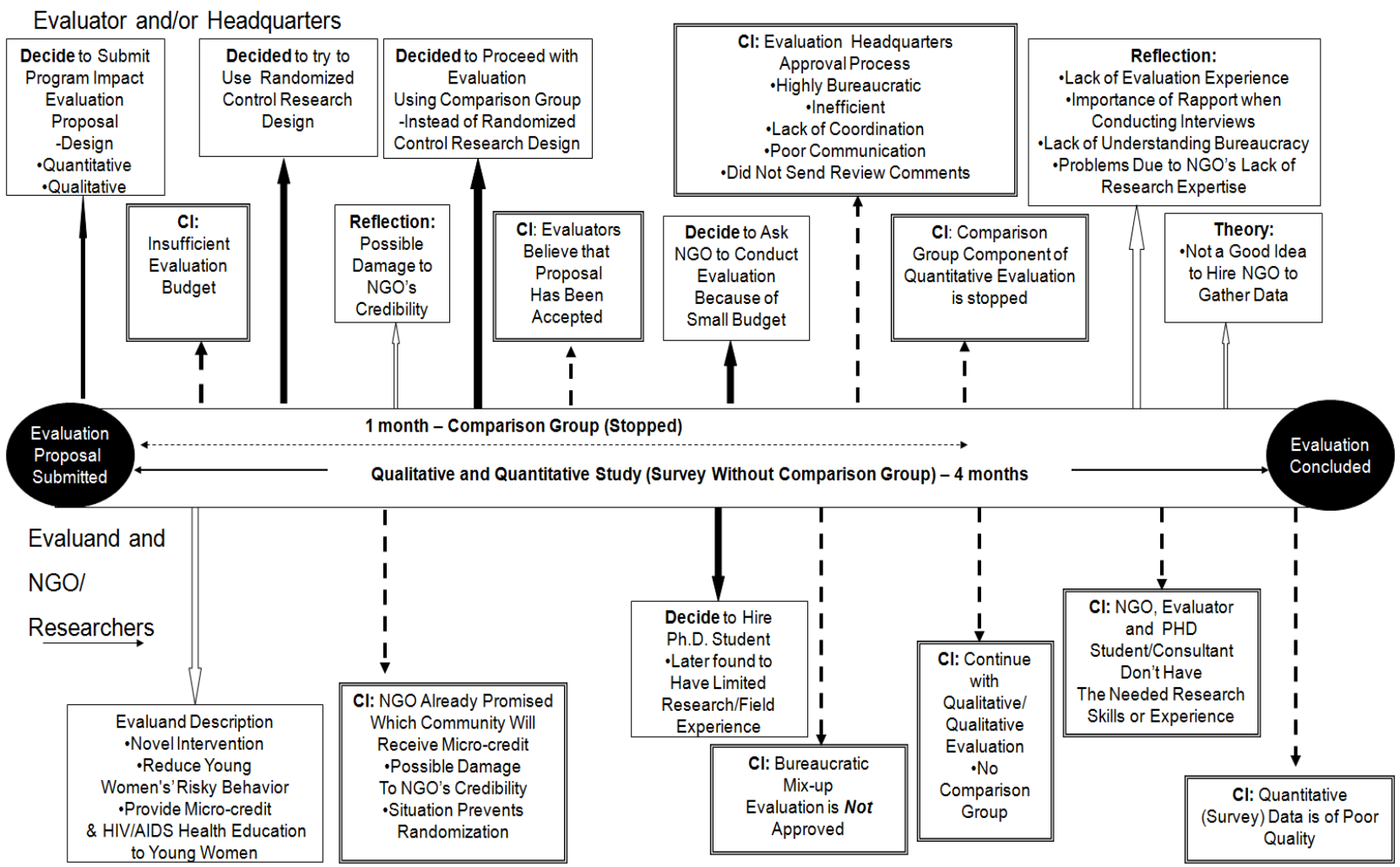

Figure D1.5. Case 3 Decision Making Timeline - Summative impact evaluation of micro-credit and an HIV/AIDS health education intervention.

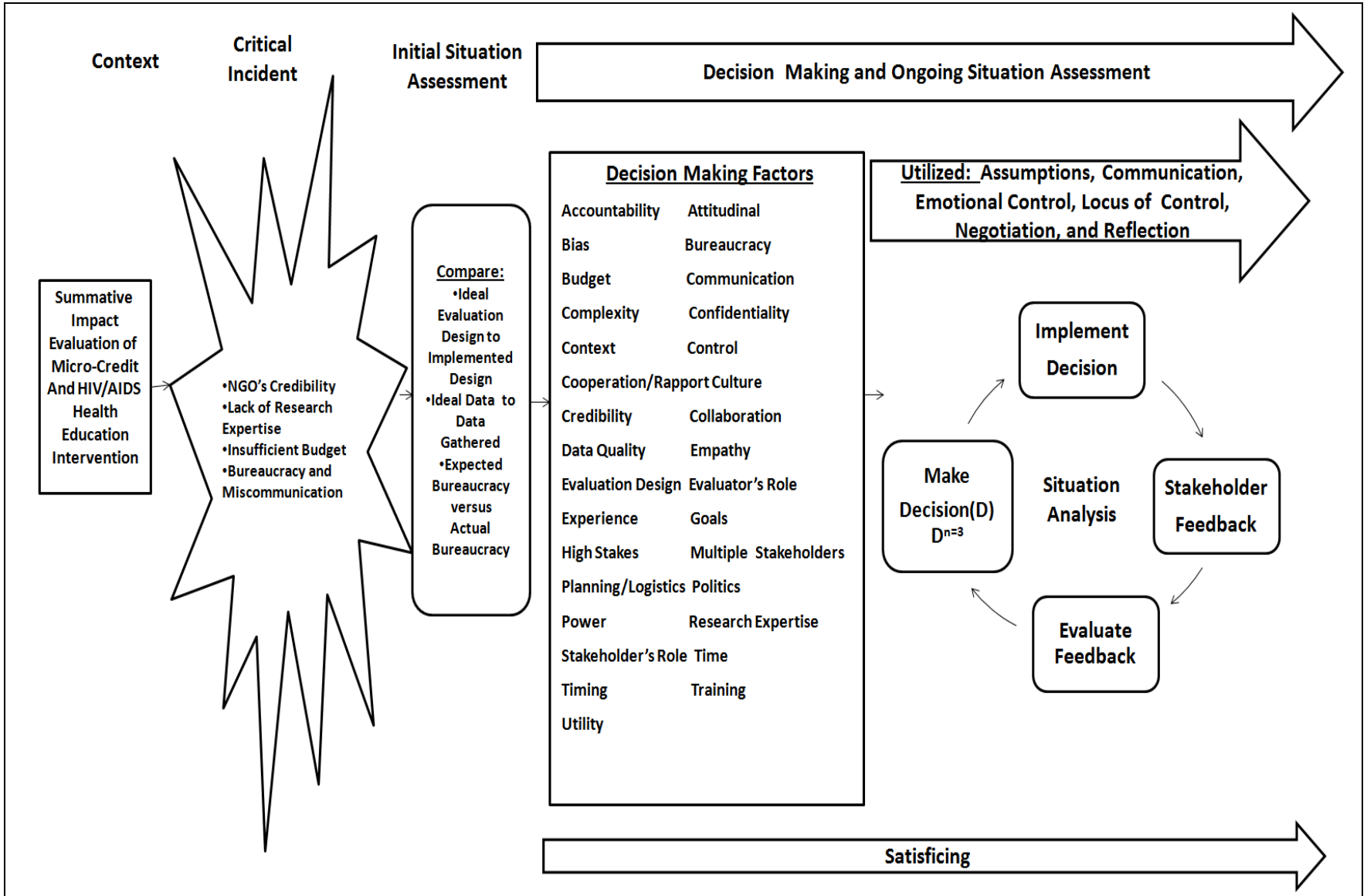

Figure D1.6. Case 3 Model - Summative impact evaluation of micro-credit and an HIV/AIDS health education intervention.

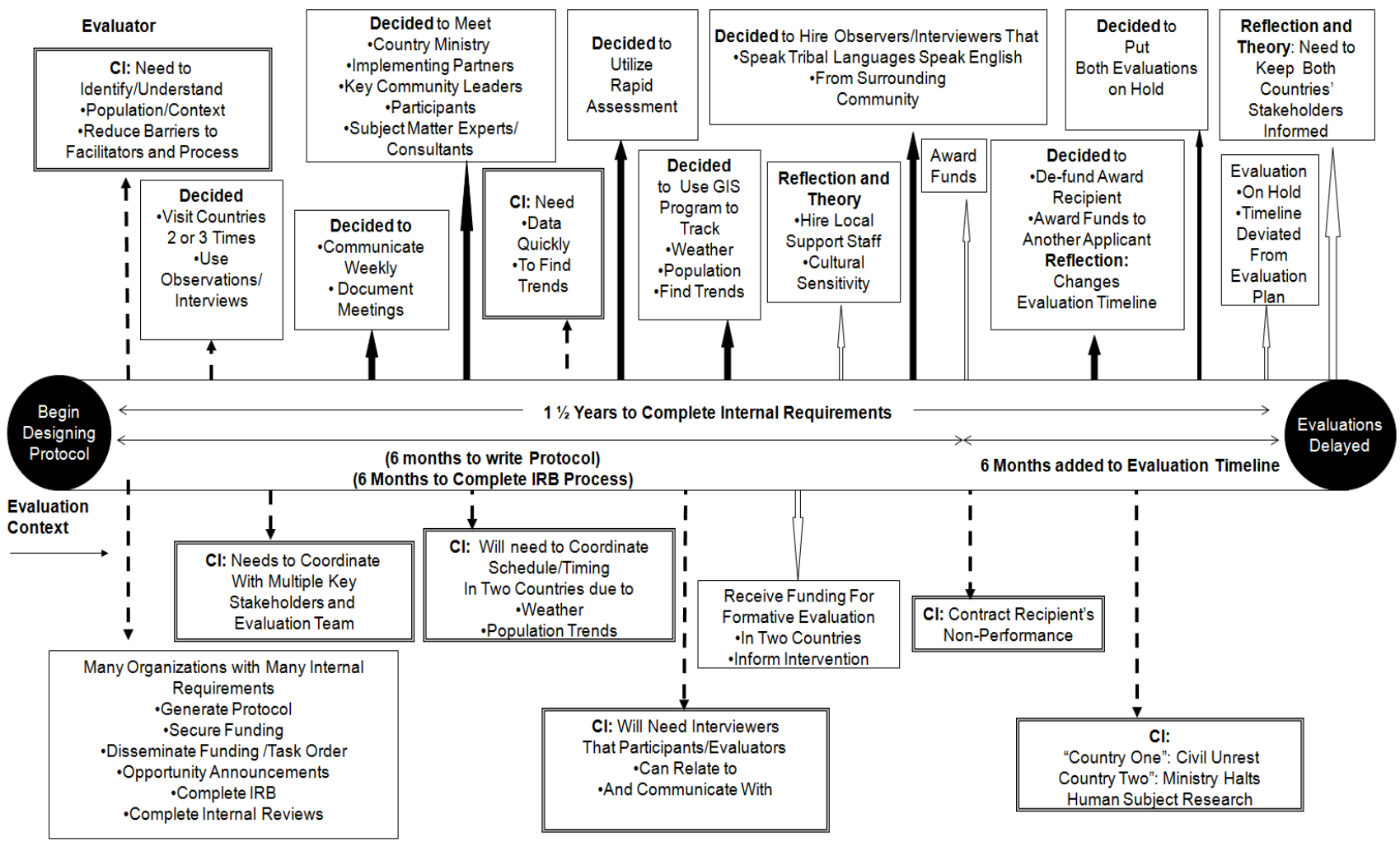

Figure D1.7. Case 4 Decision Making Timeline - Ex-Ante evaluation that was aimed at designing a transactional sex and women's HIV/AIDS vulnerability evaluation.

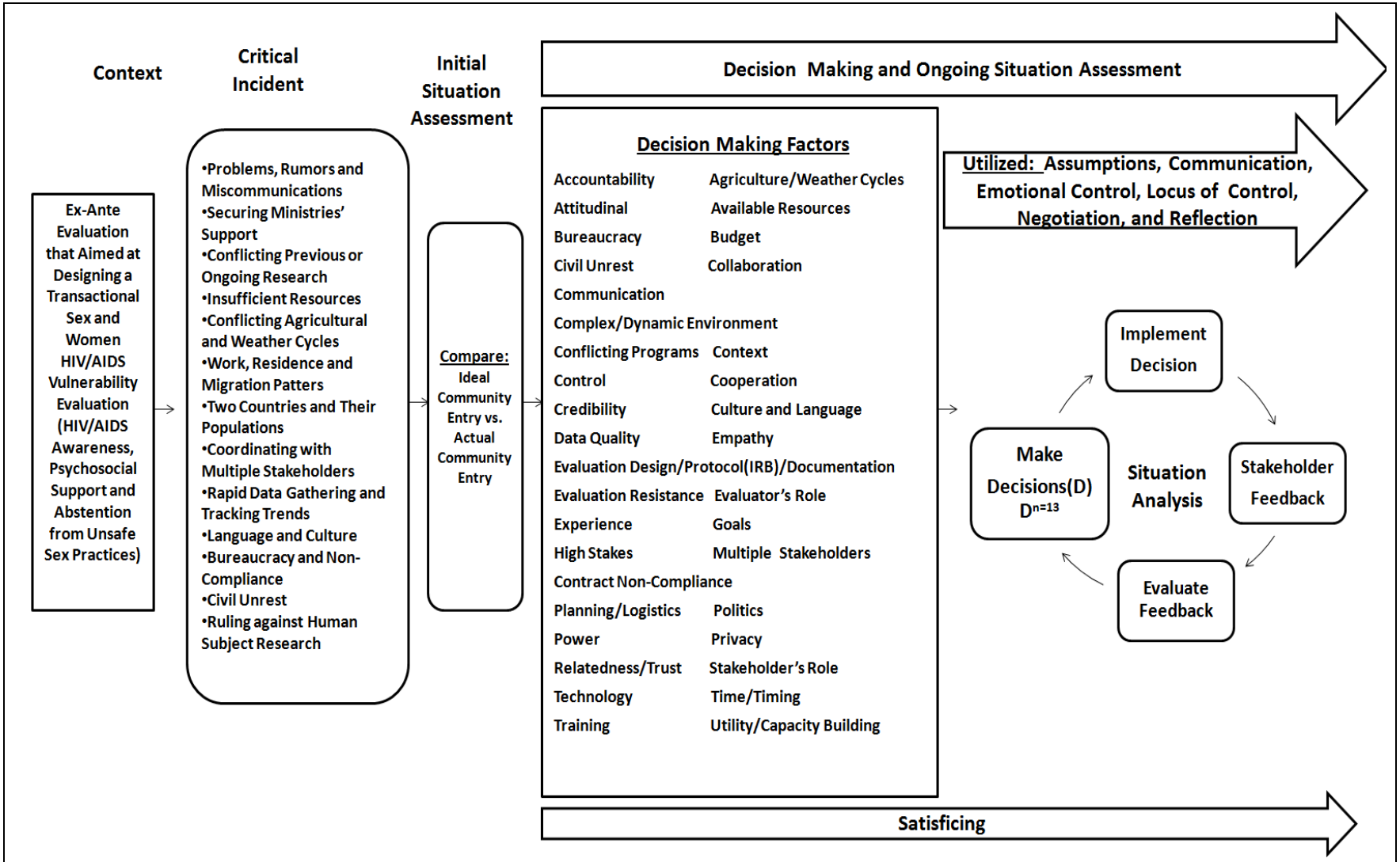

Figure D1.8. Case 4 Model - Ex-Ante evaluation that was aimed at designing a transactional sex and women's HIV/AIDS vulnerability evaluation.

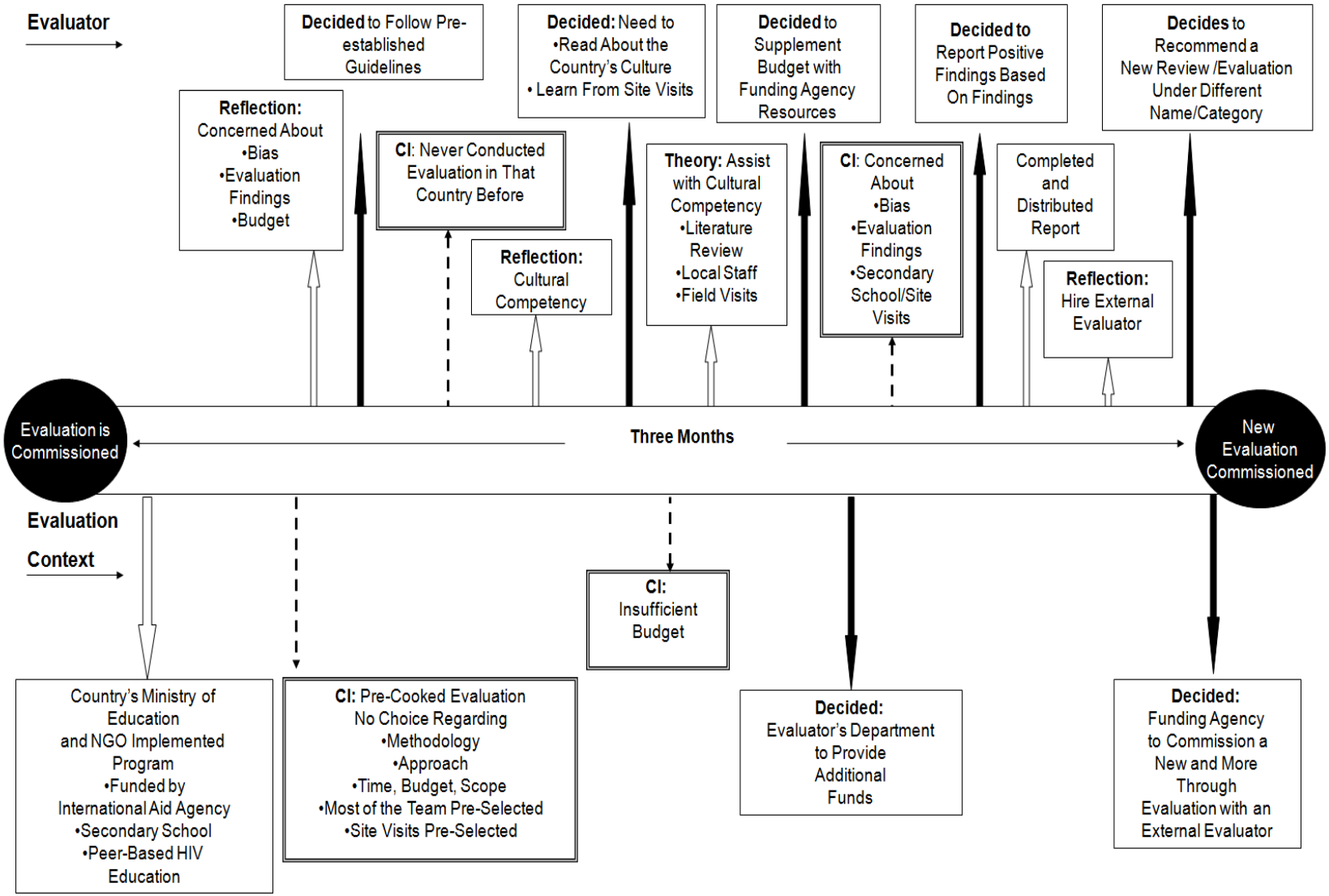

Figure D1.9. Case 5 Decision Making Timeline - Summative evaluation of a secondary school HIV health education intervention.

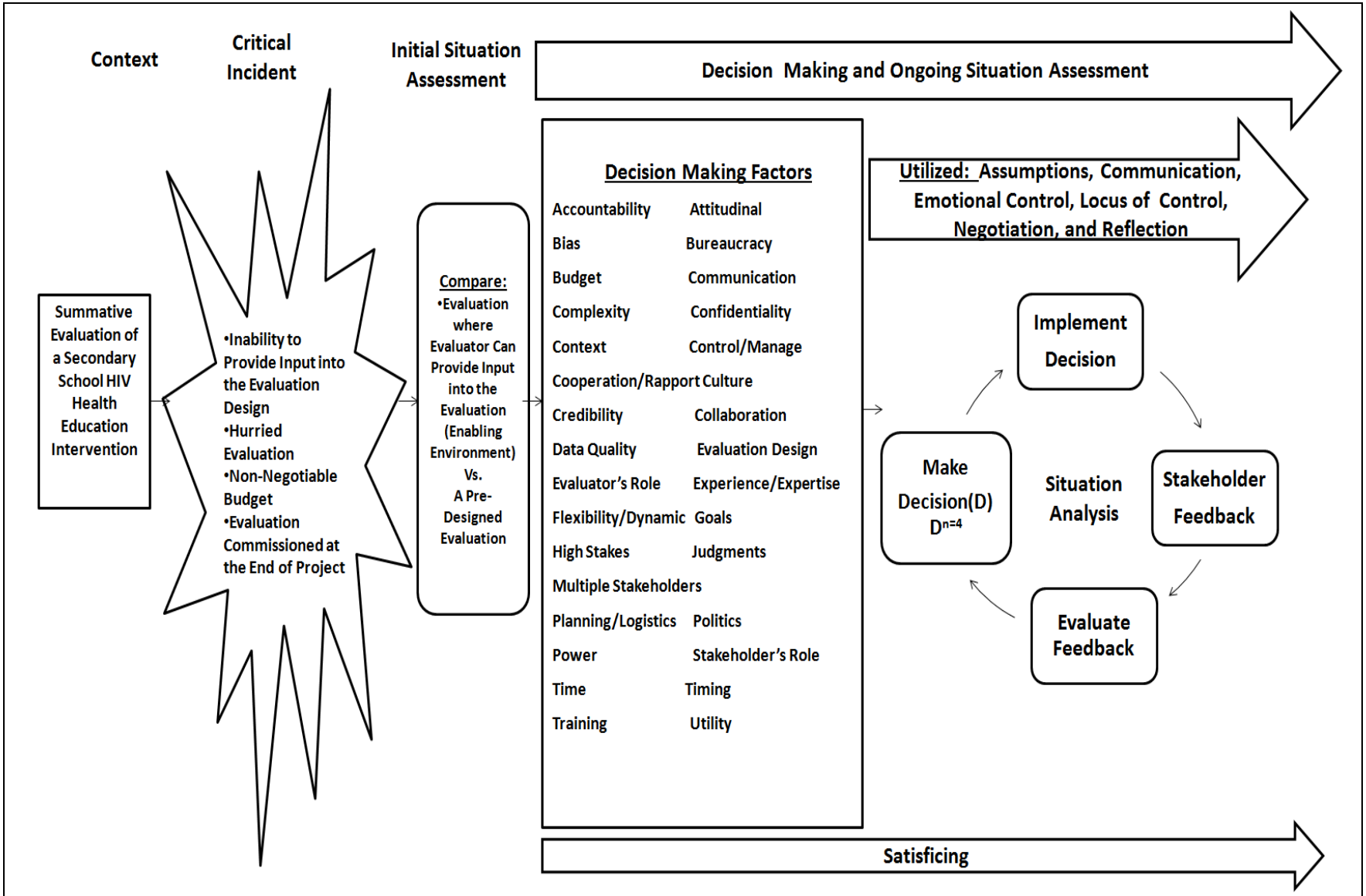

Figure D1.10. Case 5 Model - Summative evaluation of a secondary school HIV health education intervention.

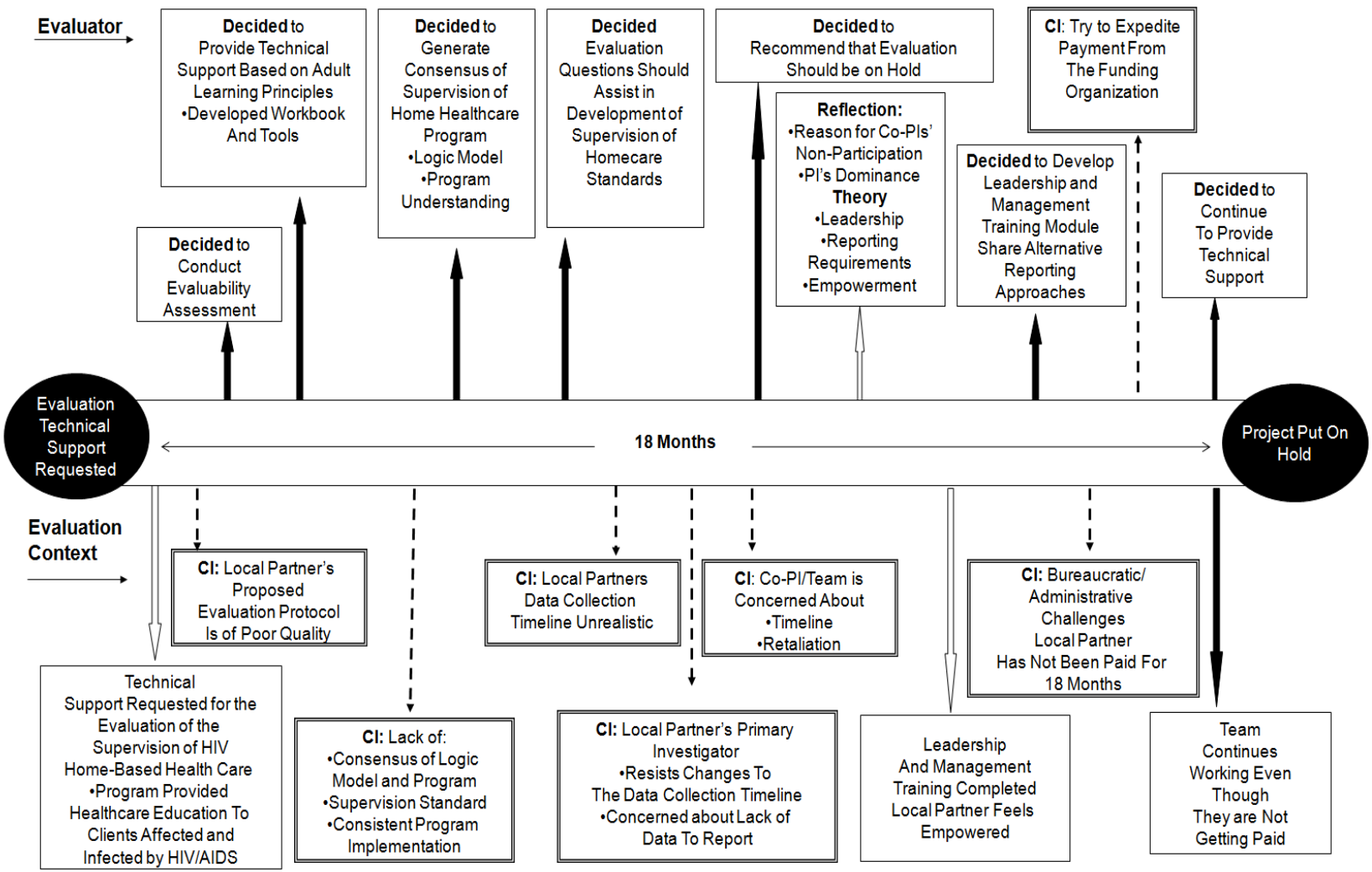

Figure D1.11. Case 6 Decision Making Timeline - Design and development of the formative evaluation of a country's HIV/AIDS home-based healthcare supervision programs.

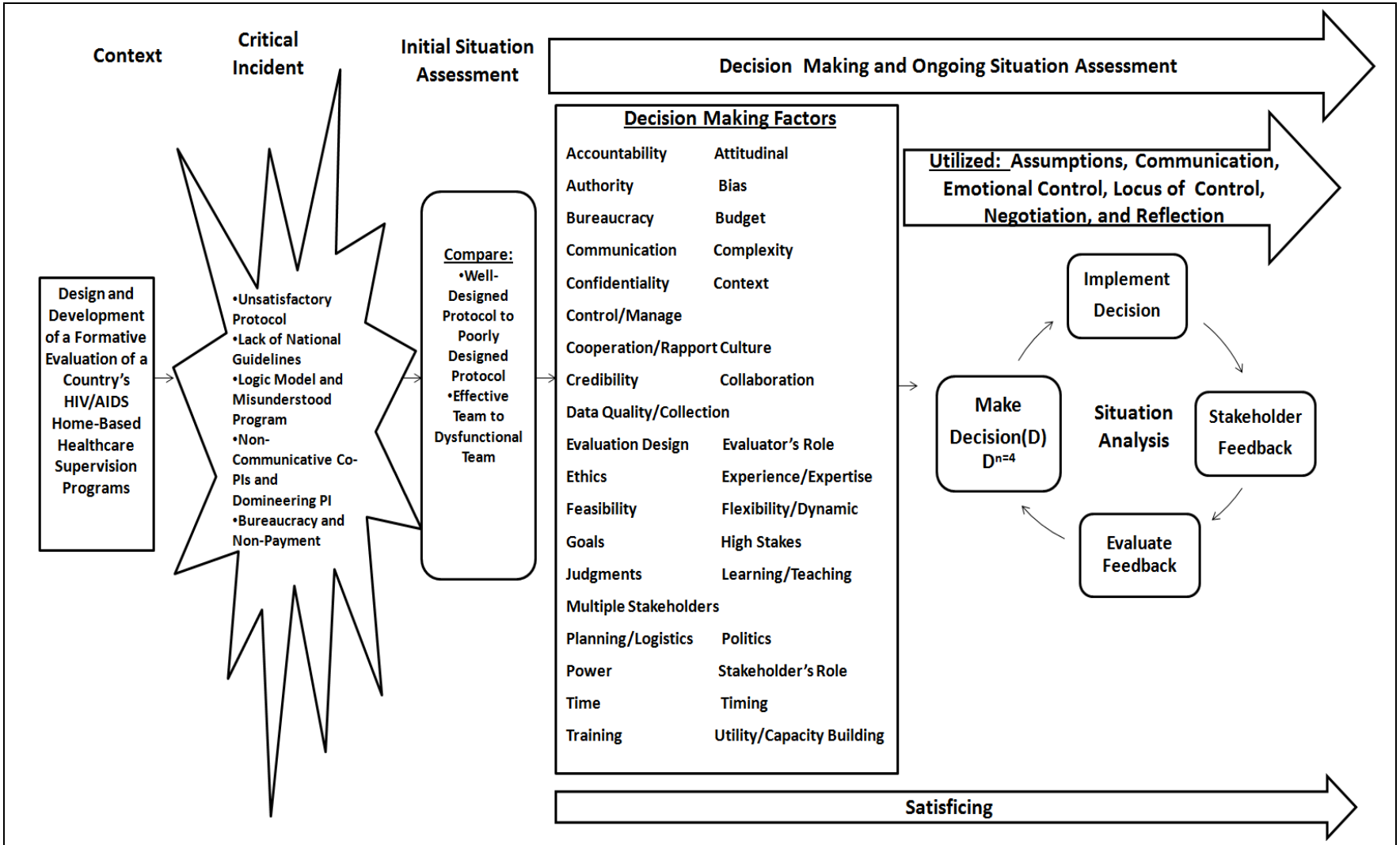

Figure D1.12. Case 6 Model - Design and development of the formative evaluation of a country's HIV/AIDS home-based healthcare supervision programs.

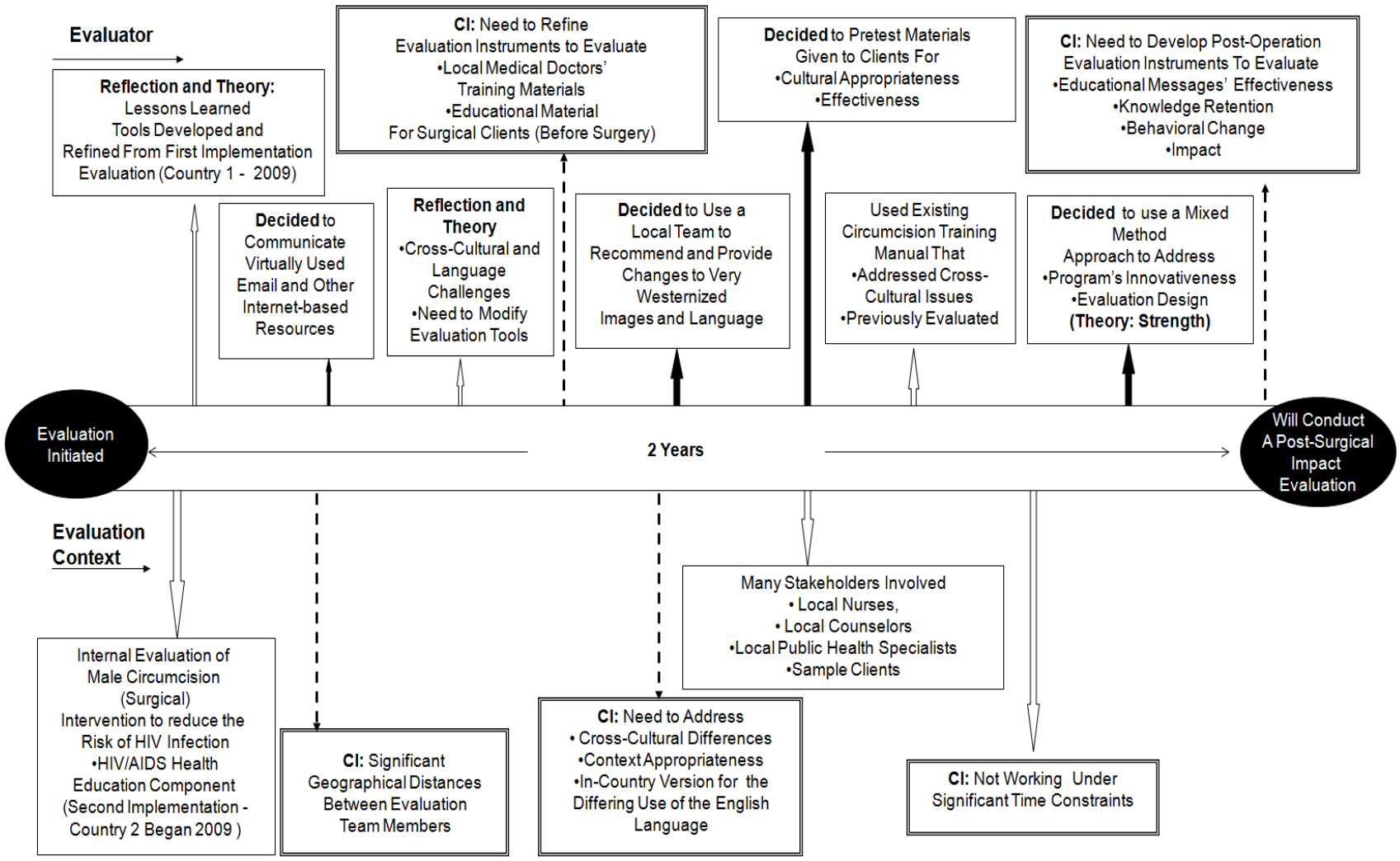

Figure D1.13. Case 7 Decision Making Timeline - Formative evaluation of the health education component of a surgical intervention that reduces HIV transmission.

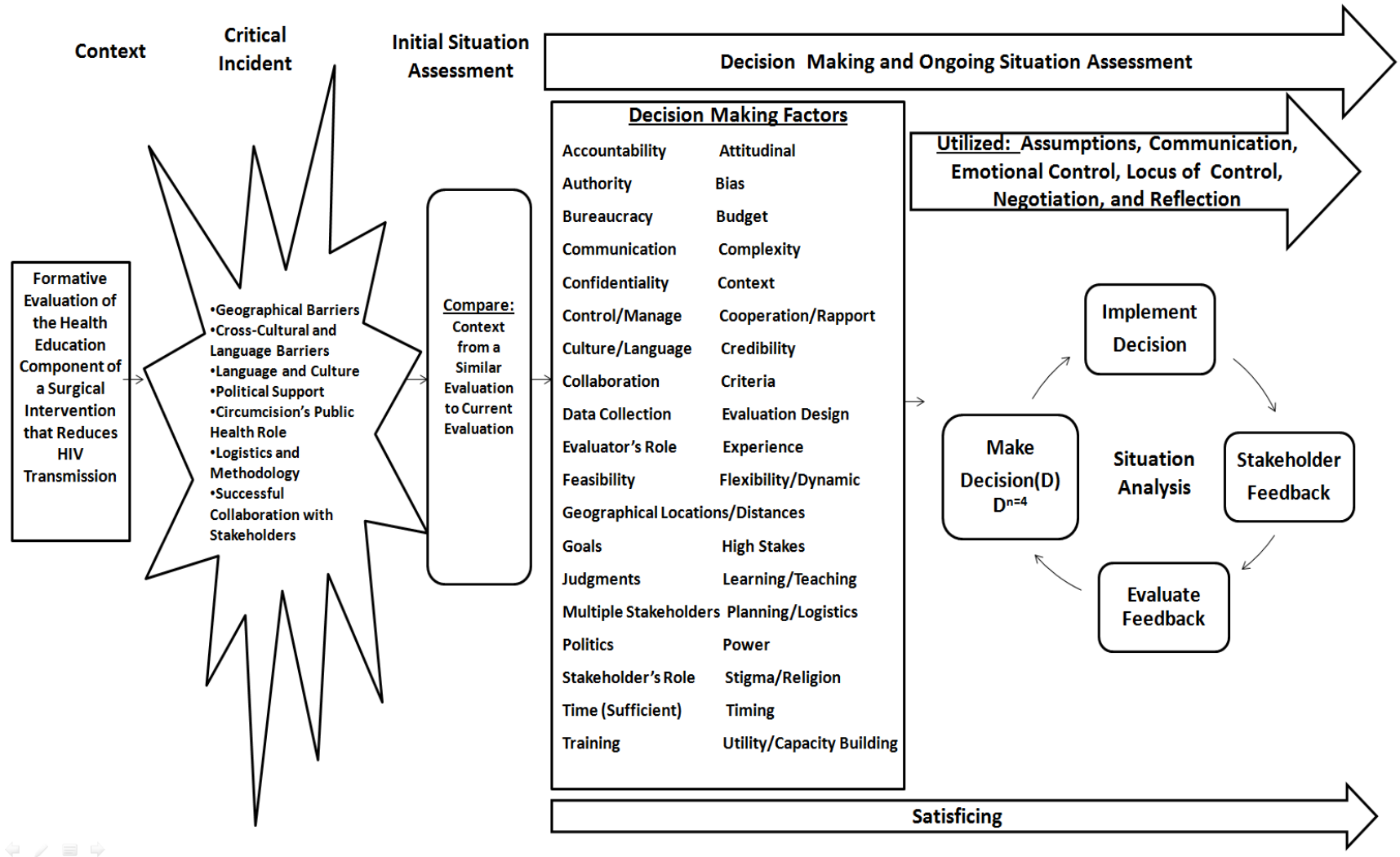

Figure D1.14. Case 7 Model - Formative evaluation of the health education component of a surgical intervention that reduces HIV transmission.