Apollo M. Nkwake

Working with Assumptions in International Development Program Evaluation

With a Foreword by Michael Bamberger

Springer

Working with Assumptions in International Development Program Evaluation

Apollo M. Nkwake

Working with Assumptions in International Development Program Evaluation

With a Foreword by Michael Bamberger

Springer

Apollo M. Nkwake
World Vision Inc.
Washington, DC
USA

Dedicated to Maureen, Theodora, Gianna and Benita

Making assumptions explicit is one way of identifying differences clarifying choices, and ultimately fostering debate and cooperation among people who are committed in some way to building a better world[1]

Pearce 2000: 40 in Eade 2003: xi

Approaches to development, and the methods that flow from them, are profoundly shaped by assumptions that are made about people…. Assumptions are also made about processes, such as how change happens or how learning takes place. Assumptions are made about what can and cannot be done. All of these shape the nature of the approach and the choice of methods. Where do these assumptions come from? Some are based on experience or sound research and evidence from elsewhere. Others are based on beliefs and values—some of which can be based on stereotypes and misinformation[2]

Rowland 2003: 4

---

[1] Eade, D. (2003). *Development methods and approaches: critical reflections. A development in practice reader.* London: Oxfam GB.

[2] Rowlands, J. (2003). Beyond the comfort zone: some issues, questions, and challenges in thinking about development approaches and methods. In D. Eade (ed.), *Development methods and approaches: Critical reflections. A development in practice reader* (pp. 1–20). Oxfam GB, London.

Foreword

All development programs and policies are based on a wide range of assumptions. These include: why (for example) combating teenage drug addiction is considered to be more important for this agency than, for example, the low level of writing skills among high-school graduates; what are the causes and consequences of the problem being addressed; what is the most effective way to attack the problem; what are the particular skills or resources that this agency can draw on; and, how will success be defined?

The evaluations of these development programs and policies are also based on a wide range of interlocking assumptions. The evaluators may have assumptions about: the real purpose of the program (which may be different from what is stated in the documents given to the evaluator); what the program is likely to achieve and what problems are likely to arise; the real motivation for commissioning the evaluation and how the results will be used (and perhaps misused). The evaluator's personal values and political orientation may also lead to assumptions about, for example, whether this is a worthwhile program, whether it is likely to have negative consequences for certain groups of women, ethnic minorities; or other vulnerable groups; and what are the real purposes of the agencies supporting the program. Additional sets of assumptions also relate to the evaluation methodology: what is the "best" or the appropriate methodology to be used; what are the appropriate output, outcome and (perhaps) impact indicators to measure; do we need to base the evaluation on a program theory; does it make sense to think about causality. Evaluators also align themselves on the quantitative/mixed methods/ qualitative evaluation spectrum with all of the assumptions that these positions entail.

The agencies commissioning and using the evaluations also have their own assumptions about what is an evaluation, what can it achieve, why it is being commissioned; and how it will be used. Some might assume that evaluators should be skilled scientists who can provide precise statistical answers to questions such as "Did the program work?", and "how much quantitative difference did it make to the lives of the intended beneficiaries?" While others might assume that evaluators should be management consultants who can provide useful guidance on

how to improve program performance. Clients may also have assumptions about the role of evaluators: Is evaluation a service that you can shop for until you find an evaluator who will provide the answer you are looking for (and who will not criticize your program[1])? Or, are evaluators really working for the funding agency (whatever they may say about being there to help you improve your program), and are they willing to ask questions and provide information of interest to the implementing agencies and national policymakers?

Given the wide, and often crucial, sets of assumptions that underlie all development programs and their evaluations, one might have expected to find that all program documents and their corresponding evaluation designs would include a detailed statement of their underlying assumptions. One might also have expected that it would be standard practice for evaluators to discuss and clarify these assumptions before the evaluation began. Even more importantly, members of the evaluation team could be expected to discuss and reach agreement on the key assumptions underlying the proposed evaluation hypotheses and research designs. However, as Apollo Nkwake reminds us, most of these assumptions are not made explicit, and in many cases, the agencies supporting programs, the managers and staff of implementing agencies, and the evaluators are often not even aware of some of these key assumptions. Based on having had to review over 200 program evaluations during the past 14 years, he tells us that "… *nothing has stunned me like the pertinence of assumptions to evaluation viability. What a resource and a risk assumptions can be! I have found them a great resource when they are explicated- their validity not withstanding – and a great risk when they are not explicated.*"

Working with Assumptions in International Development Program Evaluation offers a timely review of the complex layers of interlinked theoretical, operational, and methodological assumptions on which both development programs and their evaluations are based. He also provides a framework for identifying and understanding the logic these multiple assumptions, and proposes guidelines for evaluating the assumptions and building them into the evaluation framework.

Nkwake argues that given the multiple contextual factors that influence how programs are designed and implemented and the complex processes of behavioral change that even seemingly "simple" projects can stimulate; most development interventions should be considered as being "complex". Consequently the first four chapters are devoted to a discussion of the attributes, design, and evaluation issues involved in complex development interventions. These chapters lay the groundwork for the later sections of the book by identifying the many layers of, frequently implicit, assumptions that are built into complex interventions and their

---

[1]   An evaluation colleague working in Russia reported that several clients were annoyed to find that evaluation reports they had commissioned, criticized their organizations. "I am not paying you money to criticize my organization" was a frequent complaint from clients who had different assumptions about the nature of evaluation and the role of the evaluator. At least they made their assumptions very explicit!

evaluation. He also presents a number of different approaches to the evaluation of complex interventions, and the different assumptions on which each is based.

Part II examines evaluation theory and assumptions. A distinction is made among social science theory, evaluation theory, evaluator's theory, and program theory each of which contains critical but frequently implicit, assumptions. Ten types of assumptions are identified and classified into three broad categories: paradigmatic, prescriptive, and causal. Each category of assumptions has different implications for a full understanding of the foundations on which an evaluation is based. Chapter Seven addresses the question why are assumptions important. He generalizes from Bonoma's (1978) assertion that "the power of an experiment is only as strong as the clarity of the basic assumptions which underlie it", and argues that the same applies to development evaluation research.

Part III presents a more in-depth discussion of the nature and importance of diagnostic, prescriptive, and transformational assumptions. While the Parts I and II have a broader theoretical orientation, Part III has a more operational focus and contains many examples of how the different assumptions actually affect the design, implementation, and interpretation of evaluations. The concluding Part IV discusses how to evaluate assumptions and to identify and explicate the assumptions. Nkwake reminds us that this can be a sensitive and challenging task as stakeholders may resent being questioned about values in which they firmly believe or about assumptions that they may feel are self-evident. Michael Patton has observed that stakeholders may also resent being put in the position of schoolchildren who have to try to guess what the evaluator or workshop facilitator has already decided are the "correct" assumptions. Workshops for uncovering the theory of change or stages of a logic model are often seen as frustrating and perhaps even a waste of time.

The book can be read both as a reference text on program design and evaluation theory, or as a practical guide on the importance of assumptions, how to define and use them, and the problems that arise when assumptions are not understood or examined.

A full understanding of the role of assumptions becomes particularly important in the rapidly evolving field of mixed methods evaluation. When evaluators share a common discipline, many foundational assumptions are shared by all evaluators and perhaps do not need to be spelled out. For example, quantitative researchers may agree on the basic assumptions underlying the use of regression analysis (e.g., assumptions about the characteristics of sample distributions and the statistical power of the test); and researchers who regularly use focus groups may share assumptions about how and when it is appropriate to use focus groups. However, anyone who reads academic journals is aware of the frequent complaints that critics, frequently from within the same discipline, have made wrong assumptions about the research purpose, design, or analysis. But it can be argued that there is a shared understanding of the foundational assumptions of their discipline, and while they often disagree on how to interpret or apply these assumptions, there is a broad agreement on the nature of the disagreement.

However, when evaluators are drawn from different disciplines, there may be fundamental, but frequently unstated, differences concerning assumptions about the nature of evidence, how hypotheses are developed (or even whether it is appropriate to formulate hypotheses), what kinds of evaluation designs are appropriate, what is considered as credible evidence, and what criteria should be used for assessing the validity of findings and conclusions. In the real world of development evaluation, the risk of misunderstanding is increased by the fact that there is often little or no time for the whole team to meet together to develop a common understanding of assumptions and methodology.

The situation is further complicated by the fact that in most mixed methods design one or other discipline is dominant, and researchers from the other discipline are often brought in after the broad framework of the evaluation has already been defined and they are often asked to collect additional data to support the dominant design. When quantitatively oriented evaluators decide to incorporate a qualitative dimension, the reasons for doing this are very different than when a qualitatively oriented evaluation team decides to incorporate a quantitative dimension, and the structure of the resulting mixed methods design tends to be quite different in each case[2]. In fact, there are many evaluations of international development programs in which the quantitative and qualitative researchers rarely meet. While the quantitative researchers are designing their sample surveys, the qualitative researchers are sent off to conduct case studies—often with only fairly general guidelines on how the cases are to be selected and what questions are to be asked. In cases such as this there is little opportunity to discuss the assumptions on which the two teams base their evaluations. Even when there is closer cooperation it is often the case that one team is expected to adapt to the research paradigm defined by the dominant discipline and there may be little opportunity or interest in developing a shared understanding. This is of course only one scenario, although it is unfortunately quite common in the field of international development evaluation; and there are many examples where mixed methods evaluations are conducted with more generous budgets and less time constraints and where they are able to achieve a higher level of integration of the different approaches. However, even in the most favorable circumstances, the management of mixed methods evaluations requires a more proactive management style (Bamberger forthcoming)[3] in which additional time and resources are required to develop an integrated research team. Understanding the assumptions from which different members of a mixed methods evaluation team are starting is a challenging task, and one of the

---

[2] See Bamberger, Rugh and Mabry (2012) *RealWorld Evaluation: Working under budget, time, data and political constraints*, Chap. 14, pp. 320–324. For an example of how a quantitative dominant and qualitative dominant evaluation of a rural health program might approach a mixed methods design.

[3] "Introduction to Mixed Methods in Impact Evaluation" (scheduled for publication in 2012 in the InterAction Guidelines on impact evaluation series. This discusses the special management challenges for NGOs, as well as other kinds of organizations, in effectively using mixed methods evaluation approaches.

areas where the final five chapters of the present publication can potentially make one of its most important contributions.

Part III proposes strategies for identifying and understanding the different kinds of program assumptions on which evaluators base their approaches to evaluate design as well as the specific tools and techniques used during each stage of the evaluation. A useful distinction is made between *diagnostic* assumptions about the causes of the problem the program is addressing, prescriptive assumptions about the appropriate interventions or strategies to address the problem and achieve program objectives, and *transformational* assumption about how to reach broader, long-term goals. Chapters 8–10 describe and illustrate the different sets of assumptions (explicit and implicit) that evaluators can hold with respect to each of these three areas; and the classifications presented could provide a useful framework to use in team-building sessions to help each member understand the similarities and differences in the key assumptions of each team member. Once the different assumptions and perspectives have been brought out in the open, this can provide a starting point for either moving toward a common understanding and approach, or at the very least providing a much clearer understanding of the differences. The systematic approach presented in these chapters will provide a helpful framework for evaluators and evaluation managers to unpack the different kinds of assumptions and to understand the differences in how they are used by members of the mixed methods evaluation team.

<div style="text-align: right">Michael Bamberger</div>

Preface

Writing a book on assumptions in development program evaluation has been my interest for a long time. My inspiration comes from my experiences as a program evaluator and many conversations with other evaluators interested in this topic.

In the last 14 years, I have been involved (partially or fully) with and reviewed more than 200 program evaluations and assessments. Yet nothing has stunned me like the pertinence of assumptions to evaluation viability. What a resource and a risk assumptions can be! I have found them a great resource when they are explicated—their validity not withstanding—and a great risk when they are not explicated. One of my most vivid experiences was in 2009 when I traveled to rural Mozambique to review a community development program. I learned from my discussions with stakeholders that the program, in part, was intended to boost farmers' incomes by distributing an agricultural bulletin to them. I thought about this for a while; distributing an agricultural bulletin to boost farmers' incomes? The discussion rolled on with several questions: What would happen when the farmers receive the bulletin? How certain were we that they would read the bulletin? If they read the bulletins, what would happen? How certain were we that they would acquire the knowledge disseminated in the bulletin? If they did, what would happen? How certain were we that they would act upon it? Why would or wouldn't they?

Program staff had good answers for these questions, but it was absolutely clear to me that this was the first time the questions were being discussed openly. It also turned out that most assumptions that had been made were not valid. Yet all it took to test them was simply to verbalize them. This proved extremely useful for me and for them in assessing the viability of this particular program strategy.

Over time and from many such experiences, I have developed an interest in exploring ways in which the necessity of explicating program assumptions can be communicated to people who design, fund, implement, and evaluate development programs. This way, programs would function a lot better; stakeholders would learn a lot more from evaluations, and beneficiaries might truly be better off.

I have received much encouragement in writing this book from conversations I have had with several other evaluators interested in this topic, including Francesca

May 2012   Apollo M. Nkwake

Contents

Part IV   Conclusions

Chapter 1
Introduction

Abstract The key discussions in this book are outlined. The first two parts of the book are more theoretical, intended to review the literature on program evaluation themes most closely related to assumptions. The last two parts of the book focus on more practical discussions on how to explicate and evaluate program assumptions.

Challenges of a Growing Profession

I do not always find it easy to explain my job to people of other professions. It is not that I do not know or I do not want to say what I do. It is just that when I tell them whatever I tell them, they do not seem to get it. An immigration official asked what I did for a living as I was returning to the U.S. recently. "I evaluate development programs," I responded. The official replied, "So your company builds roads in communities and your work is to check how well it has happened." That is a bit off target, but a lot closer than many guesses I have heard over the years. My wife never has to explain what she does when she tells people that she is a pharmacist. But she once told me, "People ask me what you do and I do not know what to tell them". She knows what I do. She just finds the same problem I find in explaining what "it" actually is.

I suspect that many evaluators have encountered this challenge. In January 2012, evaluator Nora Murphy asked the evaluation community to share its "elevator speeches." It is a well-known exercise: sum up in a minute what you do to earn your living. More than 30 evaluators responded; many agreed that they struggle with explaining evaluation. One evaluator distributed basic evaluation books to family and friends. It seemed to help; she now gets fewer questions about her work. Some other elevator speeches:

I work with people who are passionate about what they do to help assess three things: Did we do the right things? Did we do things right? What could we do better?

I work with people who collect the data that informs the decisions that make the world a better place.

…The profession of an evaluator is thus revealing various expectations, collecting multi-dimensional evidences, and then compare and contrast the similarity and differences for the intended uses.

That final example may represent part of the communications gap. Even my wife might have to think twice to get the point of that statement.

While the views in this discussion cannot be generalized, they do show that responding evaluators are passionate about their work and see their jobs as important. It should be frustrating indeed for such passionate professions to find that people around them may not know so much about their discipline. Fortunately, most evaluators keep moving forward and doing good jobs. And it should not seem that their profession is a recent one.

Evaluation is not a young discipline. Evaluation practice can be traced as far back as the late 1940s and early 1950s, mostly within such U.S.-based organizations as the World Bank, United Nations, and U.S. Agency for International Development. In the early days, the focus was more on appraisal than evaluation (Cracknel 1988 in Segone 2006, p. 9). By 1986 (Scriven 1986), the discipline was speedily expanding and creating new frontiers. In 2005, Vestiman wrote of evaluation as a discipline in its adulthood.

Among the major catalysts for growth of the evaluation profession is the increased interest in and the demand for development effectiveness (Hughes and Hutchings 2011; Segone 2006). The demand for documented and preferably positive results of development activities has increased markedly over the past 30 years. These programs must demonstrate that they are not only well-intended but effective in advancing the economic, health, welfare, and/or social conditions of the recipient population.

The increasing importance of demonstrated effectiveness is documented in the Paris Declaration for Development Effectiveness, hosted by the French government on 2 March 2005 and signed by more than one hundred donor and developing countries. Parties resolved to reform aid to increase its effectiveness in combating global poverty. One of its principles, "Managing for Results", focuses on the obligation of both donor and recipient countries to ensure that resource management and program decision-making will focus on whether these produce the expected results. Donors agreed to support fully developing countries' efforts to implement performance assessment frameworks to help track progress toward key development goals (OECD 2005). This is just one example of how development actors are concerned not just about effectiveness but measurement and documentation of that effectiveness.

Measuring the level and nature of effectiveness is increasingly understood as an essential part of being effective in both current and future efforts. It is little wonder

that such mantras as "you cannot manage what you cannot measure" and "what gets measured gets done" continue to intrigue us. It is also true that sometimes "effectiveness" and "the measurement of effectiveness" have been mistaken for synonyms. As Forss, Marra and Schwartz (2011, p. 6) noted:

> … Accountability fever, results-based management and the evidence.-based policy movement contribute to a sense that everything can and should be evaluated. In many jurisdictions, evaluation is a standard operating procedure—something that should please evaluators as there is an abundance of work to be done…

Recognizing that evaluation is a critical element of identifying and measuring development effectiveness (or the lack thereof), there has been much investment in development evaluation (Bamberger 2006). There is much work to do; there are many questions about how best to do it. Debates abound about what methods and tools are best to use in what situations and when best to use them. There is almost always tension between what stakeholders want to know about development programs and what evaluations can tell them within the project's funding, design, and time constraints.

Assumptions in Responding to Evaluation's Challenges

How have evaluators addressed these challenges? Sometimes they have adjusted their tools and methods to make them better suited to unexpected complications. Sometimes the response has been to measure what is easy (or possible) to measure. Patton (1997, 2010) has illustrated this behavior very cleverly:

Someone was found looking for his watch under the light. He was asked where he could have dropped it. "In the grass," he replied. He was then asked why he was not looking for the watch in the grass. He responded, "It is because there is no light there." Often when we need to look for answers in dark (complex) places, we look where the light is, instead of getting a flashlight to create light where we need it. Perhaps we can always find something where the light is, even though it is not what we are looking for. The logical action, fetching a flashlight to look for the watch, frequently is off the table. The option is all too often considered to be constrained or ruled out all together by shortage of resources, lack of time, and other contextual circumstances.

What if the watch were dropped only where there is light? It would be much easier to find. Dropping the watch in a lighted area may feel like a stretch of this analogy, but programs are designed with some (and generally a lot of) intentionality. Their goals, objectives, and time frames (short, intermediate, or long term) provide some structure and organization. But in all too many cases, the design leaves much of the ground to be covered "in the dark".

Weiss (1995) argued that a major reason complex programs are so difficult to evaluate is that the assumptions that inspire them are poorly articulated. Stakeholders are often unclear about how the change process will unfold. In western

Uganda, a mother taught her young daughter never to eat food in the saucepan, but rather first to put it on a plate. To ensure that the daughter obeyed, the mother told her that if she ever did otherwise, her stomach would bulge. The daughter kept this in mind. One day as the two visited a local health center, they sat next to a pregnant woman in the waiting room. The girl pointed at the pregnant woman's bulging belly, announcing, "I know what you did!" The pregnant woman was not pleased; the girl's mother was embarrassed, and the girl was puzzled as to why she was getting stares from the adults in the room.

The problem was that the pregnant woman and the girl had two different theories about the cause of bulging stomachs. The same dilemma surfaces in evaluating development programs, especially complex programs. All too often, an evaluation has many stakeholders with widely varied perspectives. Each may have a completely different theory about how the program intervention(s) will cause the desired change (or even whatever particular change that stakeholder seeks). There is seldom a process for helping all stakeholders clarify their theories of change and their expectations. These are seldom made explicit. In turn, critical information is not captured, information that could help explain why the interventions succeeded or failed—and some stakeholders are left with a sense that the intervention failed because it did not meet their personal expectations. Stakeholders involved in program design and operation are often unclear about how the change process will unfold. They often glaze over the early and midterm changes that are needed in order to cause longer term changes (Weiss 1995). This lack of clarity about the interim steps needed toward the eventual outcome makes the task of evaluating a complex initiative challenging and reduces the likelihood of a useful outcome in understanding program successes as well as failures. Equally important, it reduces the likelihood of successful program replication.

Focus of this Book

Building on the increasing volume of work on theory-driven evaluation, this book discusses the crucial place that assumptions hold in conceptualizing, implementing, and evaluating development programs and provides simple ways of (a) helping stakeholders make explicit the assumptions that are being made by stakeholders and evaluators about program theory and environmental conditions and (b) enabling evaluators to develop and carry out effective program monitoring and evaluation in light of those assumptions.

The concept of assumptions is a sizable but relatively unexplored field. In no way do I pretend to discuss it comprehensively. We live a great deal of our lives based on assumptions, and we do not behave any differently in designing or implementing development programs. This book attempts to demonstrate how vital it is to recognize when we make assumptions and what assumptions we make. When we plan, we should not assume that we have not assumed.

The first six chapters of this book set the context for discussing tacit program assumptions: the fact the development programs are complex, often with many types of complexity (Chap. 2); that there are many program frameworks for designing programs in complex contexts (Chap. 3) and that there are many methodologies for evaluating complex programs (Chap. 4). Chapter 5 locates assumptions within theory. The building blocks for theory—(including a program's theory) are concepts, and assumptions are the glue that holds the blocks together. In Chap. 6, assumptions are defined; their major roles are outlined in Chap. 7.

Assumptions may be understood as beliefs that we take for granted about how the world works. They may seem so obvious to us as not to require explication. (Brookfield 1995). Stakeholders often take for granted their beliefs or expectations about how a program should (or will) work. The assumptions that are often taken for granted are the focus of this book. They may even be tacit assumptions, unarticulated by any of the stakeholders. Chapters 8, 9, and 10 outline the three categories of tacit assumptions: prognostic, prescriptive, and transformational. These tacit assumptions need to be explicated during program design. Chapter 11 points out the need to ensure that more overtly identified assumptions are examined, refined, and assessed for both implementation and evaluation use.

The logical framework approach (Davies 2005; Ambrose and Roduner 2009; Gasper 2000) and the theories of change approach (Judge 2000; Mason and Banes 2007; Anderson 2004) call for describing program assumptions during the design phase. Other approaches explicate program assumptions during evaluation as they attempt to "reconstruct" a theory—usually for a program that has been implemented. These include among others the strategy assessment approach (Leeuw 2003); goal evaluation (Friedman et al. 2006); the Policy Scientific Approach (Leeuw 2003); the Elicitation Methodology (Leeuw 2003); as well as the logical framework approach (Gasper 2000 and theories of change approaches (Anderson 2004). The approaches suggested in this book; the alternative causes approach (for diagnostic assumptions); the alternative strategies approach (for prescriptive assumptions); and the causal linkages approach (for transformational assumptions) are used to make explicit what had been tacit assumptions before or during program design. Chapter 11 focuses on how these assumptions can be monitored and evaluated along with program outputs, outcomes, and impacts. Chapter 12 addresses the challenges of explicit assumptions different from, but just as important as, implicit assumptions.

This book assumes that evaluator roles go beyond just evaluating programs to include working with stakeholders to design, implement, and monitor programs so that they become the most productive they can be in benefitting their intended audiences.

References

Ambrose, K. & Roduner, D. (2009). *A conceptual fusion of the Logical Framework Approach and Outcome Mapping,* OM Ideas Paper No. 1

Anderson, A. (2004). *Theory of change as a tool for strategic planning: a report on early experiences*. Aspen Institute: Washington, DC.

Bamberger, M. (2006). Enhancing utilization of evaluations for evidence based policy making. In M. Segone (Eds.) *Trends in development evaluation* (pp 120–142). New York: Unicef

Brookfield, S. D. (1995). *Becoming a critically reflective teacher*. San Francisco, CA: Jossey-Bass

Davies, R. (2005). *Moving from logical to network frameworks: A modular matrix approach to representing and evaluating complex programs, in Creative Commons Attribution-NonCommercial-ShareAlike site*. Available online at http://66.102.9.104/search?q=cache:cs98bp1tuXgJ:www.mande.co.uk/docs/MMA.doc+network+software+Visualyzer&hl=en&ct=clnk&cd=2&gl=uk

Eade, D. (2003). *Development methods and approaches: critical reflections. A development in practice reader*. London: Oxfam GB.

Forss, K., Marra, M. & Schwartz, R. (2011). *Evaluating the complex: attribution, contribution, and beyond*. New Brunswick, New Jersey: Transactional Publishers

Friedman, V. J., Rothman, J., & Withers, B. (2006). The power of why: Engaging the goal paradox in program evaluation. *American Journal of Evaluation, 27*(2), 201–218.

Gasper, D. (2000). Evaluating the 'logical framework approach' towards learning-oriented development Evaluation. *Public Administration and Development, 20*(1), 17–28.

Hughes, K. & Hutchings, C. (2011). *Can we obtain the required rigor without randomization? Oxfam GB's non-experimental Global Performance Framework*. International Initiative for Impact Evaluation Working paper 13 retrieved 10/10/2011 from www.3ieimpact.org

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation, 24*(1), 5–20.

Mason, P., & Barnes, M. (2007). Constructing theories of change: Methods and sources. *Evaluation, 13*(2), 151–170.

OECD. (2005). *Aid effectiveness: Three good reasons why the paris declaration will make a difference*. (Development Co-Operation Report), Paris: (OECD)

Judge, K. (2000). Testing evaluation to the limits: The case of english health action zones'. *Journal of Health Services Research and Policy, 5*(1), 1–3.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). Sage: Thousand Oaks, Calif.

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guilford Press.

Scriven, M. (1986). New frontiers of evaluation. *American Journal of Evaluation, 7*(7), 7–39.

Segone, M. (2006). *UNICEF Regional Office for CEE/CIS and IPEN, Evaluation Working Papers ISSUE #5*: New trends in development evaluation, Switzerland

Vestman, O. K. (2005). The relationship between evaluation and politics. In M. Segone (Ed.), *Bridging the gap: The role of monitoring and evaluation in evidence-based policy making* (pp. 46–69), New York: UNICEF.

Weiss, C. H. (1995). Nothing as practical as good theory: exploring theory-based evaluation for comprehensive community initiatives for children and families. In K. Fulbright-Anderson, A. C. Kubrisch & J. P. Connell (Eds.), *New approaches to evaluating community initiatives* (vol. 2, pp. 65–92), Theory, measurement and analysis. Washington, DC: Aspen Institute.

Part I
Designing and Evaluating Complex Development Programs

Chapter 2
Attributes of Complex Development Programs

Abstract This chapter draws on complexity literature to outline attributes of complex development programs. Most attributes of complexity—such as dynamism, emergence, adaptation, nonlinearity, and scale—are likely to facilitate program success, but make it difficult to measure that success. In trying to measure success of complex programs, evaluators and program stakeholders often assume the reverse—that these programs are simple, linear, and predictable. An even greater problem is that these assumptions, huge as they are, are taken for granted (implicit); neither explicated nor tested.

The More Complex, the More Assumptions

Understanding the complexity of development programs is a prerequisite for understanding the nature of assumptions we make in evaluating such programs. This does not imply that all development programs are necessarily complex; or that they have the same levels of complexity. Many important assumptions made about development programs—intentionally or unintentionally, pertain to their complexity. The first act of assuming is to ignore, unconsciously or consciously, the complexity of programs in their design, implementation, or evaluation. Conversely, the effort to explicate assumptions about how a program works or is expected to work is in itself an acknowledgment of its complexity. It is for this reason that explicit assumptions are key to conducting useful evaluations.

Designers of development programs must acknowledge and act on the fact that such programs and the contexts in which they are implemented, like most social systems, have multiple facets that in themselves may contain further complex assumptions. Drawing on the increasing volume of work on complexity theory, there is much for evaluators to learn about the attributes of complex development programs.

Complexity Theory

Evolving from holism and Gestalt theories after World War I, and cybernetics and general systems theory and its variants after World War II, complexity theory has gained much prominence in the last few decades in both natural and social sciences (Mason 2001, 2007; Holland 1992; Lansing 2003). Complexity theory suggests an intricate world of open, dynamic systems, based on nonlinear (and linear) relationships, self-organization principles, feedback loops, strange attractors, perturbations, bifurcations, phase shifts, and sensitivity to initial and changing conditions (Johnson and Bruce 1998). Johnson (1998) uses an example of the "butterfly effect" in which a very small disturbance to a complex system can lead to major consequences that would have been unpredictable based on estimates derived from an analysis of the initial conditions. For example, a street vendor's death in Tunis sparked an uprising that swept across the Middle East, bringing the "Arab spring" with dramatic economic, political, and social consequences across the region and around the world.

Complexity owes much to general systems theory. Mason (2001: 406) points out some differences between complexity and general systems theory:

- Complexity often concerns nonlinear relationships between constantly changing entities. Systems theory, in contrast, studies static entities linked by linear relationships defined by flows and stocks (e.g., energy, information),
- The stocks and flows perspective emphasizes quantities of flow, and not necessarily their quality. The qualitative characteristic of these flows (e.g. symbolic content of communication), is sometimes a point of interest for complexity research.
- Complexity research concerns how complex behaviors evolve from relatively simple local interactions between system components over time. Systems theory in contrast offers simplification and parameterization of flows and stocks. This assumes that the system exists in equilibrium; it therefore undermines the need to examine changing relationships between and among system elements. Complexity research contends that systems have emergent or synergistic characteristics that cannot be understood without reference to subcomponent relationships. Complexity research is concerned with how systems change and evolve over time due to the interactions of their constituent parts. It does not assume a "steady state."

Mason (2007) identifies three forms of complexity: algorithmic, deterministic, and aggregate complexity.

(a) Algorithmic complexity: This form contends that complexity of a system lies in the difficulty in describing the system characteristics (Mason 2007). The difficulty in describing system characteristics may arise because system characteristics are unclear or nebulous or because there are too many. An example of a nebulous program characteristic is the long-term change in relationships among target groups, such as "empowerment" which is difficult to define or measure.

System characteristics may also be difficult to describe if they are numerous. With respect to organizations, Daft (1992) in Anderson (1999, p. 416) equates complexity with the number of activities or subsystems within the organization, noting that it can be measured along three dimensions: (1) Vertical complexity: the number of levels in an organizational hierarchy; (2) horizontal complexity—the number of job titles or departments across the organization, and (3) spatial complexity: the number of geographic locations.

(b) Deterministic complexity: This suggests that the interaction of two or more variables can create a system (or systems) that are unpredictably stable or unstable. For example, an effort to empower women in a rural community could generate such externalities as marital tensions. With respect to environments, complexity is equated with the number of different items or elements that must be dealt with simultaneously by the organization (Anderson 1999). The behavior of complex systems is also surprising and hard to predict because it is nonlinear. In nonlinear systems, intervening to change one or two parameters by small amounts can drastically change the behavior of the whole system, and the whole system can be very different from the sum of the parts. Complex systems change inputs to outputs in nonlinear ways because their components interact with one another via a web of feedback loops. Simple boxes-and-arrows causal models are inadequate for describing or representing systems with complex interconnections and feedback loops (Anderson 1999; 2004).

(c) Aggregate complexity: This concept is concerned with how individuals work in harmony to create complex systems, such as economies or ecosystems. This may be the case when several community development agencies, each working in a different sector, collectively bring about improved well-being in a community.

Cilliers (2000, p. 24) summarizes the characteristics of complex systems using seven attributes:

- A complex system consists of a large number of elements that in themselves can be simple. The elements interact dynamically by exchanging energy or information. These interactions are rich. Even if specific elements only interact with a few others, the effects of these interactions may be propagated throughout the system. The interactions are nonlinear.
- There are many direct and indirect feedback loops.
- Complex systems are open systems—they exchange energy or information with their environment—and operate at conditions far from equilibrium.
- Complex systems have memory, not located at a specific place but distributed throughout the system. Any complex system thus has a history, and the history is of cardinal importance to the behavior of the system.
- The behavior of the system is determined by the nature of the interactions, and not by what is contained within the components. Since the interactions are rich, dynamic, fed back, and, above all, nonlinear, the behavior of the system as a whole cannot be predicted from an inspection of its components. The notion of "emergence" is used to describe this feature of these systems. The presence of

emergent properties does not provide an argument against causality, only
against deterministic forms of prediction.

- Complex systems are adaptive. They can (re)organize their internal structures
  without the intervention of an external agent. Some systems may display some
  of these characteristics more prominently than others.

Anderson (1999) gives six insights about complex systems:

1. Many dynamical systems (whose state at time t determines their state at time
   t + 1) do not reach either a fixed-point or a cyclical equilibrium.
2. Processes that appear to be random may be chaotic, revolving around some
   points of influence and seldom returning to the same state.
3. The behavior of complex processes can be quite sensitive to small differences
   in initial conditions, so that two entities with very similar initial states can
   follow radically divergent paths over time. Consequently, historical accidents
   may "tip" outcomes strongly in a particular direction.
4. Complex systems resist simple reductionist analyses, because interconnections
   and feedback loops preclude holding some subsystem(s) constant in order to
   study others in isolation.
5. Complex patterns can arise from the interaction of agents that follow relatively
   simple rules. These patterns are "emergent" in the sense that new properties
   appear at each level in a hierarchy.
6. Complex systems tend to exhibit "self-organizing" behavior. Starting in a
   random state, they usually evolve toward order instead of disorder.

If the attributes of complex social systems described above can be extended to
development programs, the following could be said about the nature of develop-
ment programs:

1. They are difficult to describe, control, or predict. Surprises are common when
   implementing and evaluating programs.
2. One plus one may or may not be equal to two in such systems, or a pattern of
   results may be traced from A to B, then to E, and back to C or D or some other
   nonlinear route.
3. An outcome may be a consequence of a billion program actions that stake-
   holders may not exhaustively outline.
4. Development programs usually involve a large number of elements—many
   stakeholders, many activities; many outputs, numerous strategies, a variety of
   target beneficiaries, etc. These elements themselves are seldom easy to
   understand. Even when they are easy to understand, the fact that they are many
   (hence differences in interests) makes development programs difficult to design,
   implement, and evaluate.
5. The program elements—stakeholders, activities, the environment, etc. all
   interact in dynamic ways. This dynamic interaction is key in creating program
   effects. Hence it may not be appropriate to attribute the program effects to one
   element or the other. For example, creating better learning outcomes among
   students may require the effort of students themselves, peers, teachers, school

leaders, caregivers, and the entire community. Singling out one of these could misrepresent the whole-system dynamic.

Development Programs as Complex Systems

Complex situations challenge traditional evaluation practices (Patton 2010). "In other words, doing evaluations is more of a mess than it used to be." (Forss et al. 2011, p. 7). Complexity in development programs can be defined as situations in which how to achieve desired results is not known (high uncertainty), key stakeholders disagree about what to do and how to do it, and many factors are interacting in a dynamic environment that undermines efforts at control, making predictions and static models problematic (Patton 2010). Forss et al. (2011, p. 7) describe a complex evaluation assignment:

> …that the assignment is broad and refers to multifaceted assessments of social change, that the evaluators need a high degree of professional scientific skills, that there are stakeholder groups involved in the evaluation process with differing political interests and with different strategies to influence the process and the findings….the role of the evaluator is not always clear-cut and confined to gathering data and writing a report.

According to Simon (1996) (cited in Anderson 1999, p. 215), development programs can be seen as open systems that exchange resources with the environment and consist of interconnected components that work together. Rogers (2005) makes a distinction between complicated and complex programs. Complicatedness arises more in the number of program features: "Complicated programs may have multiple parallel or complementary causal paths, that is, alternative ways of achieving outcomes; they may be multilevel programs with local, regional, national, or international layers." (Rogers 2005:13). Complexity arises more from the dynamism of program elements: "Complex programs have at least two main features: emergent properties, where precise objectives and implementation strategies are developed during implementation as specific opportunities develop, and disproportionate relationships, where, at critical levels, a small change can make a big difference, that is, serve as a tipping point." (Rogers 2005, p. 13).

The notion of complexity has been much more advanced in information and organization sciences than in community development—which is arguably more intricate. As Holland (1992, p. 17) has argued, a finance minister cannot expect the same accuracy in asking the computer to play out the impact of a policy change as an engineer can expect in asking the computer to play out the implications of tilting an airplane wing. In reference to the field of anthropology, Lansing (2003, p. 203) observes that.

> So far, only a handful of anthropologists have taken an interest in complex systems. Yet much contemporary research on complex adaptive systems is concerned with questions that have traditionally formed the subject matter of anthropology.

Lverson (2003:36) clarifies this further when he defines simple systems as dealing "…primarily with non-human/non-social environments within which interventions can be isolated, manipulated, and measured". However, not all social systems are complex. Mayne (2011) talks of simple social systems as those that deal with human and social environments, but can be reasonably delineated and controlled to some extent. Mayne argues that single "project-level" interventions in education, social welfare, and others areas can fall into this group (2011, p. 53). On the other hand, in complex development programs "multiple and often unknown confounding variables are the norm" (Dixon 1995: cited in Mayne 2011, p. 60). Mayne (2011, p. 60) outlines the nature of interventions that characterize complex systems:

- They are embedded in social settings, and hence subject to numerous exogenous influencing factors.
- They seek to change the behavior of individuals and groups of individuals.
- They involve a variety of planned or unplanned activities or strategies.
- They are specific to the particular time and place.
- They have been in place for some time.
- They do not allow for experimenting with the implementation of the program as a whole or the other influencing factors.
- They usually are made up of a series of or multiple project-level interventions, often implemented over time.
- They may involve multiple levels of government.

It is difficult to speak of any development program as non-complex. The more such programs are evaluated, the more evaluators realize and appreciate their complexity, hence the growing interest in the subject of complexity itself. It has been argued that the emergence of complex development programs, sometimes referred to as Complex Community Initiatives (CCI), can be attributed to the convergence of several trends: (a) the drive to achieve certainty of success in improving the well-being of communities and sustaining that well-being; and (b) the popularity of public–private partnerships and local action as alternatives to "big government" in dealing with social problems (Connell et al. 1995).

According to Connell et al. (1995, p. 3), the attributes of CCIs that make them particularly difficult to evaluate include horizontal complexity, vertical complexity, the importance of context, the flexible and evolving nature of the interventions, the breadth of the range of outcomes being pursued, and the absence of appropriate control groups for comparison purposes.

(a) Horizontal complexity implies that complex development programs work across a spectrum of sectors. It is easier to track results in each individual sector. However, tracking results by sector, though easy to do, undermines the role of inter-sectoral synergy. On the other hand, it is very challenging to measure the effect of inter-sectoral synergy.

(b) Vertical complexity describes the characteristic of development programs that are seeking change at the individual, family, and wider community level. Does change trickle from the individual to the family and then to the community or from the community through family to the individuals? These pathways of change are neither defined nor empirically tested.

(c) Contextual complexity concerns the many factors that are often beyond the control of stakeholders, such as macro-economic trends and political, demographic, and social factors that affect the success of development programs.

(d) Flexible and revolving interventions have developed from efforts to make programs relevant to the communities where they are implemented. Stakeholders are pressured to (and may be self-motivated to) make these programs more flexible. Hence the programs are constantly changing to adapt to changes in the communities. This makes it difficult to track progress, as targeted outcomes and indicators keep changing from what was originally conceptualized in program design documents and what may have been tracked initially.

(e) A broad range of outcomes, vital as they may be, confound measurement and tracking. The domains in which development programs seek improvements, such as community empowerment, child well-being, and local economic structures are difficult to operationalize and measure.

(f) Absence of comparison or control groups (particularly at the community level) can make evaluation difficult. As most of these programs aim to benefit all members of the community directly or indirectly, it is difficult to find "comparison" communities not benefiting from the interventions. It is thus hard for evaluators to say that changes in the target group(s) resulted from the specific intervention(s). Sometimes the social systems (e.g., national education, health and agricultural systems) create spillovers to communities that were not included in the interventions. For example, an education program implemented in a select number of schools might spill benefits to non-intervention schools where the school system is managed in clusters by the state and non-governmental organizations or the church, depending on how each system deploys and equips head teachers and teachers.

Other Characteristics of Complex Development Programs

Nonlinearity

A linear system exhibits strict proportionality with parts that are interchangeable. Cause relates to effect in a direct clear and proportional fashion. Nonlinearity implies that small actions can produce large reactions (Patton 2010). It also implies that different parts of a larger whole are connected to each other and that they are not connected in a direct fashion (Forss et al. 2011, p. 11). Uphoff (1992) in Forss et al. (2011, p. 11) outlines some properties of nonlinear change:

> The distance between cause and effect can be long in time and sometimes short, and whichever depends on a large number of intervening factors; there are usually several causes for any change that occurs, and causality must be understood as multiple at best; there is no proportional relationship between the size of causes and the effects. Tiny causes may lead to larger effects, and large causal factors sometimes have limited effects; change occurs through qualitative leaps and bounds, and is neither incremental nor proportional. Boundary conditions play a major role in explaining how change occurs. Feedback loops can be positive (amplifying deviation from an equilibrium state) or negative (reducing deviation, bringing the system back toward the original equilibrium state).

Forss et al. (2011) provide an example of program nonlinearity as the asymmetric relationship between program and management on the one hand, and impact and results on the other hand. Change may be fast or slow. This may be the case for a behavior change program that is intended to provide information to people e.g., mothers—on how to feed young children. The provision of information does not guarantee that mothers will change feeding practices.

Dynamism

Complex social programs are dynamic. According to Pawson et al. (2011, p. 519), programs are active, not passive. Patton (2010) refers to this characteristic as emergence, in which patterns emerge from self-organization among interacting agents, interacting elements, and agents' response and adaptation to each other.

Pawson et al. (2011) identify six elements of dynamism:

(a) Interventions do not work in and of themselves; they only have effect through the reasoning and reactions of their recipients. Consider, for example, that a service would be delivered according to customers' or beneficiaries' feedback on the quality of service and their satisfaction.
(b) Programs have long implementation chains and multiple stakeholders. Recipients are many and varied; reactions to programs thus differ; outcomes are thus generally mixed. For example, for some programs, stakeholder participation is such a priority that how—even when—participatory program design delays implementation, participation in the design may be an achievement in itself.
(c) Programs are embedded in complex social systems. Recipients are rooted in different communities, institutions, cultures, and histories, all of which influence the success of a program. For example, a water and sanitation program may not be implemented the same way in two different communities that have the same water and sanitation problem but different cultural practices and beliefs that affect hygiene behavior.
(d) Programs are implemented amid the turbulence of other interventions. The policy agenda is delivered through a multitude of interventions, each one interfering with the reception of another.
(e) Programs beg, steal, borrow, and adapt. Practitioners work constantly to improve the delivery of interventions rather than preserving uniformity to meet evaluation and trial requirements. For example, NGOs sometimes find

themselves working in the same communities without much coordination. Community leaders may be called in for several different meetings, each about planning similar activities by different NGOs.

(f) Programs are the offspring of previous interventions. Social problems are long-standing; interventions evolve to try to combat them. The success of a current scheme depends on its history. As programs are implemented, assumptions are tested and kept or discarded for new ones, outcomes and Indicators change, and all this becomes important program history.

(g) Programs change the conditions that make them work in the first place. An intervention's success is always time limited, since alleviating a problem always involves changing its concomitant causes. One program strategy evolves into another, e.g., from relief to rehabilitation to development. Some program phases or elements may be more successful than others. Each program phase poses new challenges and requires revision of strategies.

Dynamic Stratification

Stame (2004) describes another attribute that makes programs complex—dynamic stratification. According to Stame, stratification implies that actors and constituents are multiple, and can be referred to as strata; actors are embedded in their own contexts, and each aspect that may be examined and dealt with by a program is multifaceted. Davies (2004; 2005) uses the term "scale" to refer to dynamic stratification in large non-governmental organizations.

Large non-governmental organizations typically work in several countries, and this gives them necessary diversity on the ground. However, greater scale may be associated with greater uncertainty about the effectiveness of these organizations, since it is difficult to measure performance at a continent-wide or global level. But Davies argues on the basis of Ashby's Law of Requisite Variety—that a model can only represent some aspect of reality if it has sufficient variety to capture the complexity of that reality. The wide representation (scale) may be an advantage in as far as it helps such large international organizations to be sophisticated enough to "nurture the change processes in which they are involved" (Davies 2004, p. 101).

Based on Ockham's Law of Parsimony (that there should be no more complexity than is minimally necessary), Davies also argues that although scale may be of benefit, greater scale may come with limitations, such as those attached to central planning within firms and states. Undoubtedly, the larger an organization grows, the more complex it becomes because of the larger number of players and constituents. At the same time, a larger number of actions in varied contexts imply varied approaches and styles, making both the organization and its work more complex. It could also be true, as Davies has argued, that the complex diversity within the organization equips it to respond to a diverse number of constituencies, which in the same vein enhances the organization's diversity.

Complexity and Certainty

Uncertainty adds to complexity and makes it difficult to respond (e.g., evaluate) or to act within complexity. Stacey (2002) provides a matrix for navigating through appropriate management actions in a complex adaptive system based on the degree of certainty and level of agreement among stakeholders on the issues at hand.

According to Stacey (2002), "close to certainty" issues or decisions are nearer to certainty when cause and effect linkages can be determined. This is more usually the case when a very similar issue or decision has been made in the past. One can then extrapolate from past experience to predict the outcome of an action with a good degree of certainty.

On the other end of the certainty continuum, "far from certainty" issues may be unique or at least new to stakeholders. The cause and effect linkages are not clear. Extrapolating from past experience is not a reliable method to predict outcomes in the "far from certainty" range. According to Stacy, stakeholders respond to the issues on the different levels of the certainty continuum according to the level of agreement they have on the various issues.

Conclusion: Complexity and Assumptions in Program Evaluation

All stakeholders would like programs to succeed, but success may mean different things to different people. Program success is often thought to be measured by the extent to which a program achieves intended (or unintended) results (Shao et al. 2012; Lycett et al. 2004; Reiss et al. 2006; Partington 2000; Maylor et al. 2006; Kassebaum 1990; Pellegrinelli 1997; Lycett et al. 2004). But program success is not limited to program results. It goes beyond outlining outcomes and impacts of a program to understanding why the program performed the way it did. This is especially necessary if stakeholders are to make needed adjustments in the program or carry the lessons from one setting to another (Wolman 1981; Newton and Llosa 2010). Another measure—program efficiency—addresses delivery of results within timeframes and budget, meeting functional requirements, measuring program effects on stakeholders and their satisfaction, and gauging sustainability, and broad program impacts (Shao et al. 2012). In this book, the term "program success" is used mostly in connection with program results.

Tables 2.1 and 2.2 outline several attributes of complex development programs. The question is, is the complexity of these programs a bad thing? Does it diminish or increase the chances of program success (i.e., the program achieving its goals)? And does it facilitate or hinder the measurement of program success (including other motives for program evaluation other than its success)? Table 2.1 shows that most complexity attributes such as dynamic interaction, emergence, adaptation, and scale may facilitate program success. However, most of the complexity

Table 2.1 Does complexity facilitate program success?

Complexity attribute	(a) Is this attribute likely to facilitate program success?	(b) Is this attribute likely to hinder program success?
Difficulty in describing program characteristics, e.g., pursue amorphous goals that are not easily operationalized (algorithmic complexity)	Yes: The pursuit of an amorphous high level goal like "empowered communities" may be the most appropriate response to complex social problems like "lack of self-sufficiency"	Yes, as argued in column A. However, the difficulty to define or describe program characteristics makes it difficult to devise appropriate interventions and this could compromise success
Self-organizing behavior—starting from a random state to a stable state (deterministic complexity). Can (re)organize internal structure without the intervention of an external agent (adaptation)	Yes. This implies that programs are flexible, and can adjust and manage to survive in the face of negative change	Yes, as in column A. However, flexibility could drift a program away from its primary focus (e.g., goals, target groups)
The interaction of even the seemingly simple elements creates complex systems (aggregate complexity); The nature of interaction among elements is more important than what is contained in the individual elements. The behavior of the entire system cannot be determined by the inspection of its elements (emergence)	Yes. It implies that even small interventions could produce large changes	Yes, as argued in column A. However, since "emergence" is unpredictable, the pathways of large changes from small interventions may not be defined or strengthened.
An intervention to create a small change in one of the program components could drastically change the entire system; resist simple reductionist analyses that require holding one system constant as another is studied or deterministic predictions of causality (nonlinearity)	Yes, a program could have positively surprising results	Yes, a program could have negatively surprising results

(continued)

Table 2.1 (continued)

	(a) Is this attribute likely to facilitate program success?	(b) Is this attribute likely to hinder program success?
Large number of program elements, e.g., many different stakeholders with different interests; many communities; target groups, or a variety of planned and unplanned activities; many interventions and strategies and sectors/ multiple levels of change-individual, family, community, institutions, organization, or governments, etc.	Yes. Scale increases reach	Yes, as in column A. However, scale could be a challenge for management and implementation
A large number of dynamic direct and indirect feedback loops (dynamic interaction). There are varied reactions to the program from different stakeholders, non stakeholders, other programs, etc.	Yes. Feedback, especially from stakeholders, allows appropriate reshaping of the program.	Yes, as in column A. However, it may not be easy to manage dynamic feedback that may be contradictory and confusing
Open systems-embedded in and influenced by their environments	Yes and No	Yes and No; the environment presents both opportunities and threats
Have memory/history not located in a specific place in the system	Yes and No. Programs continuously influenced by their past, positively or negatively depending on that past.	Yes and No. As in column A
Uncertainness-Cause–effect linkages are not clear	No. Since decisions are difficult to make	Yes, as in column A

Table 2.2 Does complexity facilitate measurement of program success?

Attributes of complex development programs (systems)	(a) Is this attribute likely to facilitate measurement? of program success	(b) Is this attribute likely to hinder the measurement? of program success
Measurement of program success		
Difficulty in describing program characteristics, e.g., pursue amorphous goals that are not easily operationalized (algorithmic complexity)	No, as in column B	Yes. What is difficult to define is difficult to measure
Self-organizing behavior—starting from a random state to a stable state (deterministic complexity). Can (re)organize their internal structure without the intervention of an external agent (adaptation)	No. Flexibility compromises consistence, making it difficult to compare program results at different time points	Yes, as argued in column A
The interaction of even the seemingly simple elements creates complex systems (aggregate complexity); The nature of interaction among elements is more important than what is contained in the individual elements. The behavior of the entire system cannot be determined by the inspection of its elements (emergence)	No, it is difficult to attribute "emergent" outcomes to specific program elements	Yes, as in Column A
An intervention to create a small change in one of the program components could drastically change the entire system; resist simple reductionist analyses that require holding one system constant as another is studied or deterministic predictions of causality (nonlinearity)	No. many evaluation models make deterministic assumptions	Yes, as in column A
Large number of program elements, e.g., many different stakeholders with different interests; many communities; target groups, or a variety of planned and unplanned activities; many interventions and strategies and sectors/multiple levels of change-individual, family, community, institutions, organization, or governments, etc.	No. Examining a multitude of perspectives requires a lot more expertise and resources	Yes, because of the difficulty of evaluating "synergy" of many elements or find control/comparison communities

(continued)

Table 2.2 (continued)

Attributes of complex development programs (systems)	(a) Is this attribute likely to facilitate measurement? of program success	(b) Is this attribute likely to hinder the measurement? of program success
A large number of dynamic direct and indirect feedback loops (dynamic interaction). There are varied reactions to the program from different stakeholders, non stakeholders, other programs, etc.	Yes. Evaluation is as an avenue of feedback can tap into a program's feedback mechanisms	Yes, as in column A. However, feedback needs to be triangulated to create appropriate responses to evaluation questions
Open systems-embedded in and influenced by their environments	No. The challenge to the evaluator is to study linkages from the program's environment that may influence results positively or negatively and not to take them for granted	Yes. As in column A
Have memory/history not located in a specific place in the system	Yes. The organization's or program's or communities' past is an opportunity for better understanding of program behavior	Yes. Synthesizing historic with current perspectives requires greater expertise and resources
Uncertainness-Cause–effect linkages are not clear	No. Cause effect linkages are not clear	Yes, as explained in column A

attributes, as shown in Table 2.2, simply make it difficult to evaluate program success.

Thus, complexity as a feature of development programs may be necessary for success, but it unquestionably makes it difficult for evaluation. However, we evaluate programs so that we can ultimately make them successful. We should not minimize program complexity simply for the sake of making programs measurable. Evaluators need to appreciate the complexity attributes of programs and adjust their tools and approaches accordingly. However, what stakeholders and evaluators often do not do is to reduce programs in an actual sense, but to assume this reduction or simplification. We often assume that the programs are simple, linear, certain, etc. Interventions are devised on these assumptions, and evaluators devise measures and methods for assessments, based on these assumptions. Patton (2010) refers to this as the ceteris paribus hoax. In what Patton calls a "Ceteris Paribus world", "all things are equal, or holding all else constant, in which the environment is simply assumed to be stable, constant and nonintrusive" (p. 197). According to Patton, "such an assumption makes for nice, neat, bounded, controlled and fundamentally misleading evaluation studies, if the object of study…just happens to be taking place in the real world" (Patton 2010, p. 197).

Certainly these programs happen to exist in the real world where the ceteris paribus assumption is inoperative. There are two dangers in the ceteris pluribus hoax. The first is that making such assumptions is unrealistic, coupled with the fact that complexity is too serious to take for granted. This is dangerous enough, but there is a worse danger: the fact that these assumptions, as stakeholders make them, are not explicated, and evaluators are often not aware that they have been made. The evaluator's assumption is that these assumptions have not been made, in the same manner that those making the assumptions assume that "everyone knows" what is being assumed. The unearthing of "taken for granted" assumptions needs to be a priority not only when programs are being evaluated, but much earlier, when they are designed. Chapter 3 appraises some approaches commonly used to design programs and how they can help in the unearthing of "taken for granted" assumptions. Chapter 4 suggests a rule of thumb for selecting methodology(ies) in evaluating complex programs.

References

Anderson, P. (1999). Complexity theory and organization science. *Organization Science, 10*(3), 216–232.

Anderson, A. (2004). *Theory of change as a tool for strategic planning: A report on early experiences*. Aspen Institute: Washington D.C.

Cilliers, P. (2000). Rules and complex systems. *Emergence, 2*(3), 40–50.

Connell, J., Kubisch, A. C., Schorr, L. B., & Weiss, C. H. (1995). *New approaches to evaluating community initiatives: Concepts, methods, and contexts*. Washington, D.C: Aspen Institute.

Davies, R. (2004). Scale, complexity and the representation of theories of change. *Evaluation, 10*(1), 101–121.

Davies, R. (2005). *Moving from logical to network frameworks: A modular matrix approach to representing and evaluating complex programs in creative commons attribution-non commercial-ShareAlike site*. Retrieved [http://66.102.9.104/search?q=cache:cs98bp1tuXgJ: www.mande.co.uk/docs/MMA.doc+network+software+Visualyzer&hl=en&ct=clnk&cd= 2&gl=UK].

Forss, K., Marra, M., & Schwartz, R. (2011). *Evaluating the complex: Attribution, contribution, and beyond*. New Brunswick: Transactional Publishers.

Holland, J. H. (1992). Complex adaptive systems. *Daedulus, 121*(1), 17–30.

Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, *21*, 93–110.

Johnson, J. E. V., & Bruce, A. C. (1998). Risk strategy under task complexity: A multivariate analysis of behavior in a naturalistic setting. *Journal of Behavioral Decision Making, 11*(1), 1–17.

Kassebaum, D. G. (1990). The measurement of outcomes in the assessment of educational program effectiveness. *Academic Medicine, 65*(5), 293–296.

Lansing, J. S. (2003). Complex adaptive systems. *Annual Review of Anthropology, 32*, 183–204.

Lverson, A. (2003). *Literature review: Attribution & aid evaluation in international development*. Ottawa: IDRC Evaluation.

Lycett, M., Rassau, A., & Danson, J. (2004). Program management: A critical review. *International Journal of Project Management, 22*(4), 289–299.

Mason, S. M. (2001). Simplifying complexity: A review of complexity theory. *Geoforum, 32*, 405–414.

Mason, R. B. (2007). The external environment's effect on management and strategy: A complexity theory approach. *Management Decision, 45*(1), 10–28.

Maylor, H., Brady, T., Cooke-Davies, T., & Hodgson, D. (2006). From projectification to programmification. *International Journal of Project Management, 24*(8), 663–674.

Mayne, J. (2011) Contribution analysis: Addressing cause effect. In K.Forss, M. Marra & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution, and beyond* (pp. 53–96). New Brunswick, New Jersey: Transactional Publishers

Newton, A. X. & Llosa, L. (2010) Toward a more nuanced approach to program effectiveness assessment: Hierarchical linear models in K–12 program evaluation. *American Journal of Evaluation*, *31*(2), 162–179

Partington, D. (2000). Implementing strategy through programmes of projects. In J. R. Turner & S. J. Simister (Eds.), *Gower handbook of project management* (3rd ed., pp. 33–46). Aldershot: Gower.

Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guilford Press.

Pawson, R., Wong, G., & Owen, L. (2011). Known knowns, known unknowns, unknown unknowns: The predicament of evidence-based policy. *American Journal of Evaluation, 32*(4), 518–546.

Pellegrinelli, S. (1997). Programme management: Organizing project based change. *International Journal of Project Management, 15*(3), 141–149.

Reiss, G., Anthony, M., Chapman, J., Leigh, G., Payne, A., & Rayner, P. (2006). *Gower handbook of programme management*. Aldershot: Gower.

Rogers, P. (2005). Evaluating complicated—and complex—programs using theory of change. *Evaluation Exchange, 11*(2), 13.

Shao, J., Müller, R., & Turner, J. R. (2012). Measuring program success. *Project Management Journal, 43*(1), 37–49.

Stacey, R. D. (2002). *Strategic management and organizational dynamics: the challenge of complexity* (3rd ed.). Harlow: Prentice Hall.

Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation, 10*(1), 58–76.

Wolman, H. (1981). The determinants of program success and failure. *Journal of Public Policy, 1*(4), 433–464.

Chapter 3
Designing Complex Development Programs

Abstract Approaches to designing development programs vary. Some emphasize stakeholder involvement; others focus on a program's environment; still others emphasize the sequence of change depicted in program results (change frameworks). This chapter examines the use of such change frameworks as the Logical framework approach (LFA), Theory of change approach (ToCA), and Participatory impact pathways approach (PIPA) for designing complex development programs and the extent to which such frameworks enable stakeholders to explicate and question implicit assumptions.

Introduction

Program design (or planning) involves the selection of an integrated set of purposeful activities that will be undertaken to achieve one or several related objectives. Decisions made during this process can be political in the sense that they involve compromises and tradeoffs among various interests, though the differences may be couched in technical terms (Abonyi and Howard 1980). Here the term program design includes issues that have to be decided prior to selection of activities and problem definition. One of the crucial aspects of designing complex programs, as Chap. 2 suggests, is unearthing the assumptions that shape the program designs. Whatever process stakeholders may take in program design or in devising a plan to implement the design, unearthing the assumptions underlying these processes is vital to ensuring sound program structure and to uncloaking potential problems. Surfacing "taken-for-granted" assumptions must be high on the list; failure to make these assumptions explicit could be fatal to program success.

Program design approaches can be grouped into three categories:

(a) Stakeholder-focused approaches: These emphasize why and how to engage optimal involvement of stakeholders, such as beneficiaries and program sponsors, in designing programs. Hummelbrunner (2010) refers to these approaches as social frameworks.
(b) Change-perspective: These encompass perspective(s) that can be applied to any kind of change framework, especially systems thinking.
(c) Outcome-focused approaches: These concentrate on program outcomes as the framework for their planning (including simple linear processes like logical framework approaches and complex reiterated processes).

The first two—stakeholder-focused approaches and change perspectives are described briefly below. The remaining discussion in this chapter chiefly addresses outcome-focused approaches, with particular emphasis on the LFA.

Stakeholder-Focused Approaches

Stakeholder-focused frameworks tend to emphasize the roles of various actors in the change process. An example is Social network analysis (SNA), a set of techniques for analyzing social systems, used to understand networks and their participants. SNA helps program designers to visualize, measure and simulate relationships (See Owen-Smith and Powell 2004; Ahuja 2000; Moody and White 2003; Chandrasekhar et al. 2011; Freeman 2006; Wasserman and Faust 1994; Hogan et al. 2007). SNA also allows designers to analyze these relationships in visual, as well as mathematical terms (Hummelbrunner 2010). Programs involving multiple actors and operating at different levels of scale can benefit substantially from the SNA technique.

Another framework that emphasizes the role of stakeholders in a change process is Outcome mapping (OM). Developed by the International Development Research Centre (IDRC) in Ottawa, Canada, OM guides program stakeholders through an iterative process to identify their desired change(s) and to work collaboratively to bring it (them) about. It focuses on changes in the behaviors, relationships, and actions or activities of the people, groups, and organizations with whom a development program is working directly or seeking to influence—and of the program(s) being influenced by these interactions (See Deprez and Ongevalle 2006a, b; Earl et al. 2001; Kibel 2000; Rockwell and Bennett 2000). OM focuses on one specific type of outcome: changes in the behavior of 'boundary partners' with whom the program works directly and anticipates opportunities for mutual interaction and influence (Hummelbrunner 2010).

Change Perspective

A change perspective is not a change framework per se; it is a way of thinking that can be applied to several different frameworks. A common perspective is "systems thinking," an approach that has been popularized by Williams and others (e.g., Williams and Iman 2008; Trochim et al. 2006; Leischow et al. 2008). Johnson et al. (2011) argue that systems thinking is a form of evaluative thinking that involves mining "systems literature for its heuristic value" as opposed to using systems methodologies (e.g., system dynamics, network analysis) in evaluation work. Richard (2009) argues that systems concepts (such as interrelationships, system dynamics, perspectives, boundaries, etc.) can be particularly useful to evaluators in situations where rigorous rethinking, reframing, and unpacking of complex realities and assumptions are needed. Applying systems concepts in designing programs, according to Richard, yields several benefits:

- more effective ways of dealing with complexity and complex situations.
- development of new ways of understanding situations.
- enabling of attention to coalitions by project evaluators.
- focus on properties that emerge unexpectedly.
- help in identifying leverage points.
- capacity to enable stakeholders to recognize evolutionary nature of programs.

Outcome-Focused Frameworks

These frameworks focus on the sequence of changes that are anticipated to be created by the implementation of a program. Davies refers to these as "theories of change." (Davies 2004:103) describes five categories of change-focused processes:

- linear processes with varying numbers of stages, e.g., logical frameworks.
- linear processes with branching structures (e.g., when revisions are made on a logical frameworks to include more steps, outputs, and outcomes).
- simple Parallel Processes (nested Logical Frameworks): output of a program's Logical Framework can be represented as the 'purposes' in component projects. Differences in the approach are then documented in each project, at their output levels.
- interacting Parallel Processes: data on activities and outcomes are aggregated into larger units, such as localities or responsible staff, rather than longer term outcomes in the lives of beneficiaries.
- reiterated Processes: These create linkages between goals at different levels in the hierarchical structure (projects and objectives).

The present discussion focuses on the logical frameworks approach, which has been the most commonly applied to program planning and evaluation.

The LFA and Assumptions: Does It Adequately Reflect Complexity?

In the last 5 decades, the LFA has come to play a central role in the planning and management of development (e.g. Ambrose and Roduner 2009; Bakewell and Garbutt 2005; Davies 2005; Gasper 1997; 2000, etc.). With its origins in United States military, space and development agencies, the LFA was later adopted and adapted by European development organizations and United Nations agencies. It has become a major standard in applications to donors for grants.

Its distinguishing feature is the use of the four-by-four matrix–the log frame. A log frame is a summarized description of a design of an intervention or program showing how inputs will be transformed into outputs and outcomes. But there is more to the LFA approach than the log frame. The matrix simply summarizes the main elements of the project and connects them to each other. The LFA is an overarching process by which the elements that go into the matrix are formulated—the wider planning procedures of problem analysis, the development of objectives and indicators, and identification of risks and assumption, which then feed into the overall program plan (Bakewell and Garbutt 2005). LFA is a widely used (sometimes misused) model for planning; its roles will be explored in subsequent sections and it will also serve as a point of reference in discussing other models.

The log frame has a reasonably standard form. A typical example of the format is given in Table 3.1. The content of the four-by-four matrix is straightforward: (i) the vertical column is a hierarchy of objectives: *activities* deliver *outputs* that contribute to *outcomes*, which help to bring about the *overall goal* (Bakewell and Garbutt 2005). In some variations, the four rows are called activities, results/outputs, purpose(s), and overall objectives (Gasper 2000, p. 18). (ii) The horizontal row demonstrates how progress against each objective can be assessed (indicators and means of verification) and the external factors (assumptions and risks) that might affect whether reaching these objectives will contribute to the next level.

The ZOPP

One offspring of the logical framework approach is ZOPP (Zielorientierte Projektplanung or GOPP—Goal-Oriented Project Planning), which has been used and promoted by GTZ (German Technical Corporation) ZOPP is an extension of the LFA. The ZOPP's output is a planning matrix—the logical project framework—that summarizes and structures the main elements of a project and highlights logical linkages among intended inputs, planned activities, and expected results. The ZOPP approach is used for essentially all German-funded projects and is a prerequisite for funding approval. The acronym 'ZOPP' is sometimes used to encompass all of the logical framework methodologies in deference to its principal initiators, the German development agencies, and particularly the GTZ (World Bank 2009).

Table 3.1 European commission's version of the project matrix

Intervention logic	Objectively verifiable indicators	Sources of verification	Assumptions
1. Overall objective The longer term benefits to (target-group) beneficiaries and wider benefits to other groups	Indicators Measures (direct or indirect) to verify to what extent the overall objective is fulfilled	Data sources for indicators for overall objective	This cell is empty in the EC version but some versions put here: Important events, conditions, or decisions necessary for sustaining objectives in the long run
2. Project purpose Benefits to be received by the project beneficiaries or target group	Indicators Measures (direct or indirect) to verify to what extent the project purpose is fulfilled.	Data sources for indicators for project purpose	1. Assumptions Important events, conditions, or decisions outside the control of the project which must prevail for the overall objective to be attained
3. Results Services to be delivered to the intended beneficiaries or target group	Indicators Measures (direct or indirect) to verify to what extent the results are produced	Data sources for indicators for results	2. Assumptions Important events, conditions, or decisions outside the control of the project management, necessary for the achievement of the project purpose
4. Activities The activities that have to be undertaken by the project in order to produce the outputs	5. Inputs Goods and services necessary to undertake the activities	Data sources for activity monitoring	3. Assumptions Important events, conditions, or decisions outside the control of the project management, necessary for the production of the results 4. Preconditions

Adapted from European Commission (1999)

ZOPP includes various subparts used to clarify projects, and the project's logical framework itself is often required by development agencies in their project appraisal. There are five logical levels of the ZOPP in a standard project cycle (see Box 3.1).

Box 3.1 Levels of the ZOPP Cycle

1. Pre-ZOPP: an in-house exercise by agencies in preparation for a project.
2. Appraisal ZOPP: an in-house appraisal for preparing Terms of Reference of a project.
3. Partner ZOPP: prepared in-country; coordination of conclusions, and recommendations with staff of project country
4. Take-off ZOPP: prepared in-country; preparation of the plan of operations with personnel responsible for project execution counterpart authorities.
5. Replanting ZOPP: prepared in-country; adjustments during project implementation.

Other ZOPPs are recommended annually in projects to update planning as needed. The ZOPP has two phases: analysis and project planning. The analysis phase has 4 sub-steps, with the identification of 'real' problems as the driver for the exercises (World Bank 2009).

Other Variations of the LFA

Over the years, the use of the LFA has evolved to see many variations. Bakewell and Garbutt (2005) in their study on use and misuse of this approach have identified LFA as a formal procedure, a way of thinking, and a brand, as well as a system. (See Box 3.2).

Box 3.2 Variations in LFA

LFA as a Formal Procedure

For some, the LFA is a formal procedure for planning projects, and in some cases also providing the base for the monitoring and evaluation system. The LFA sets out a number of standard steps to be completed, which may include some form of participatory problem assessment and identification of aims and objectives, some form of risk assessment, and so forth. The output of the

Box 3.2 Variations in LFA   31

LFA, the project or program plan is usually required to be summarized in a logical framework.

LFA as a Way of Thinking

LFA is understood more as a set of ideas and a way of thinking. The value lies in working through a hierarchy of objectives, ensuring that you consider appropriate indicators, and take account of the risks and assumptions. How one goes about this process can be quite flexible and does not necessarily need to be captured in a matrix.

LFA as a Brand or Label

A third sense in which the term LFA is used is as a way of branding or labeling one's planning approach to make it acceptable to donors. As one NGO respondent stated after describing the organization's approach to planning, "We only call it LFA because the donor requires us to use the LFA."

LFA as a System

The major problem with the LFA is how it is used in practice. While many accepted that the logic and systematic approach it encouraged are very important, actually putting it into practice in their management systems presented major challenges.

Another variation of the logical framework is for programs to focus on mapping out program results without necessarily specifying assumptions indicators or means of verification. An example of a behavior change communication program is shown below (Fig. 3.1).

How is the LFA Helpful in Dealing with Complexity and Assumptions?

Although the logical framework has become universally known, it is far from universally liked. It has been the subject of much criticism over the years, concerning both its theoretical basis and its application in practice (Bakewell and Garbutt 2005). Gasper (2000) argues that the LFA approach has been a practice with relatively little accompanying theory, in sharp contrast to cost benefit analysis or even participatory appraisal (Gasper 2000).
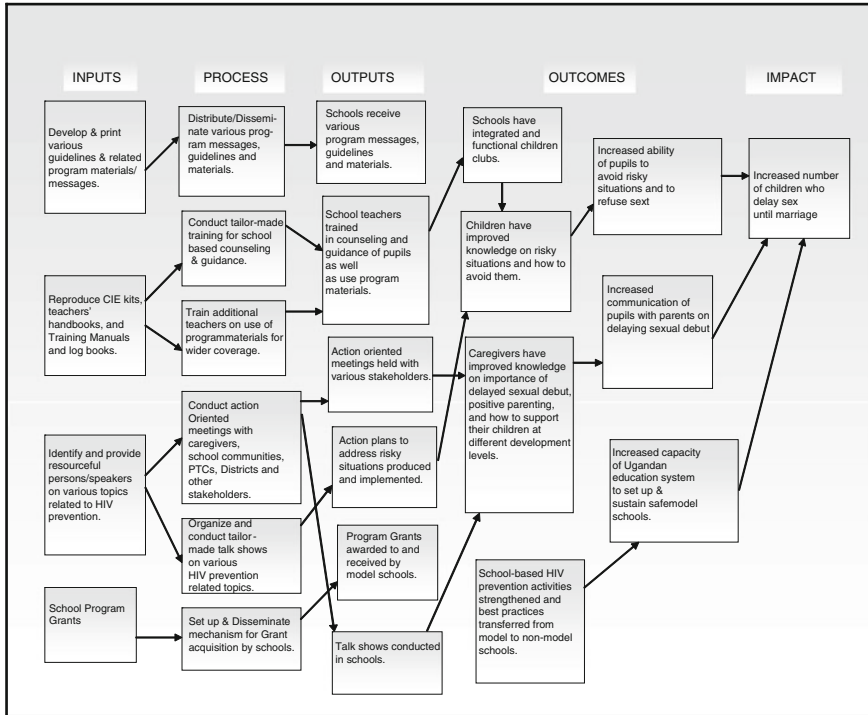
Fig. 3.1 A logic model for a behavior change communication program

One of LFA's main pitfalls lies in the identification of problems as a starting point. It starts with the perspective that the proverbial glass is half-empty rather than half-full. The 'glass half empty' perspective then drives the process, which usually results in a limited vision of potentials and solutions. Those who have worked with appreciative inquiry may not find the problem-led perspective an appropriate way to engage with communities or even to start any engagement (Gasper 2000).

Another area of criticism surrounds its flexibility. There is an understanding among users that development programs are too complex to fit in a four-by-four matrix. This pitfall is closely linked to the fact that for most people, the LFA is the matrix (and by reference the program model) and what cannot make it into the matrix is probably not worth knowing about the intervention. In struggling to describe the intervention through the vertical and horizontal logics, users find it rigid—limiting exploration, adaptive thinking, and further innovation. Program managers like summaries and perhaps this is why donors insist on the log frame. It gives them the sense of having a coherent summary of the major program elements, but it does not give them an overview of the problems, actors, assets, and linkages, to name just a few missing elements. Bakewell and Garbutt's review insists that the LFA is mostly top-down.

A simplistic characterization of the prevailing attitudes to the LFA runs as follows: "donors insist on it, while NGOs use it under sufferance. All recognize that it has many weaknesses, but there is a common view that despite these weaknesses, it is the best of a bad bunch of options available for planning and monitoring development work; Hence it carries on being widely used against all objections" (Bakewell and Garbutt 2005, p. 1).

Lessons from the LFA

Its pitfall notwithstanding, LFA has taught—or should have taught—two important lessons to development students and practitioners, including evaluators. First, the log frame gets program planners to think carefully through what they are planning to do. Planners have to consider systematically how their proposed activities might contribute to the desired goal through delivering outputs and outcomes. This has been a key in engaging clarity on steps that comprise program activities. This clarity, however, comes with simplification. The second lesson concerns the explication of assumptions. These two lessons are discussed below.

Simplification or an Extension of Clarity

Much criticism regarding the LFA centers around its tendency to simplify complex conditions and programs. Although these criticisms have basis, it is worth noting that "simplification" is not a new notion. For a long time in development and management science, people have used various tools to simplify complex situations, not as a way of overlooking their complexity but as a key step in understanding the puzzle. Examples of other such tools include maps, dashboards, and similar devices.

The example of dashboards is worth exploring a little further: A dashboard in analysis and planning, like the dashboard in an automobile, is meant to organize and present information in a way that is easy to read and use. This inevitably involves a simplification. The fact that the dashboard does not provide ALL the information does not mean that everything that does not appear on the dashboard is not useful or does not exist.

Bob Williams (2011:1) narrates an interesting experience with perceptions of the log frames. Facilitating a workshop on log frames, Bob started by putting a toy elephant on the table and asked participants to tell him what it was. Most people said that it was an elephant. So he asked what an elephant does. Among other responses, participants replied, "It provides transportation." "It moves logs." "It performs tricks in circuses." He asked, "Does this elephant do this in any conceivable way?" His students acknowledged, "No … it's a model of an elephant." The model does not provide a simplified form of reality (e.g., moves only tiny logs very short distances); it merely represents reality. It's a metaphor for an elephant.

The model at best reminds us what we've heard elephants might do, maybe even seen them doing, but it doesn't actually reflect that behavior."

It can be argued that a greater form of simplification would be to assume that we can and should understand a complex world without any aid that initially guides us to its core elements. Although simplification can aid understanding and help to focus on the most important pieces of the puzzle, there is a danger of overlooking other pieces outside the "dashboard". An example of this is the tendency to overlook or ignore unexpected results. Other examples of this nature of log frame misuse are cited by Gasper (2000, p. 21).

- A logic less frame: where a pre-existing log frame format is used to accommodate a pre-existing design rather than to help create a logical design in an appropriate format. The failing can be counteracted; a log frame format can be used from an early stage in project design, with its series of means-ends links derived from a systematic prior cause and effect analysis.
- Lack frame: Many LFA users have underestimated that a 'frame' includes some things and leaves others out, and that a 'frame-work' is to help the required work, not substitute for it.
- Lock frame: After an LFA has been prepared; it tends to be fixed and not updated, and thus becomes a lock-frame. Ironically, this seems more likely the more care has gone into an LF and the more people consulted. In principle, this failing can be counteracted if staff are permitted, willing and able to do the extra work of updating.

Perhaps the simplification pitfall in the use of log frame arises most frequently when users take it to be more than a simplification and abstraction. In doing this, the results chain turns out to be a mechanistic set of causes and effects operating in a self-contained predictable system and environment.

As Rowland puts it (2003, p. 5)

> …the tool as a whole has attracted much criticism over the years, but this seems to be one of its strength: to have a tool which systematically encourages you to become aware of the assumptions being made throughout a planning process seems like an excellent idea.

The existence of the assumptions column, however, cannot substitute for the awareness and understanding of the assumptions that are needed in order to fill it in that requires an openness of mind which the tool itself cannot provide. It also requires a willingness to revise assumptions when they prove inaccurate and on the part of institutions, a willingness not only to allow revisions, but to welcome them as evidence of learning and experience. The value of a log frame thus varies from one use to another. Program design frameworks such as the theory of change approach discussed later in this chapter emphasize the need for clarity on how program resources are expected to transform into desired results. The difference is that the LFA does the same with a matrix, which makes it appear linear. More often than not, this is a false depiction of the transformation process. A more appropriate representation of the "messiness" of the complex transformation processes is shown in Fig. 3.2.
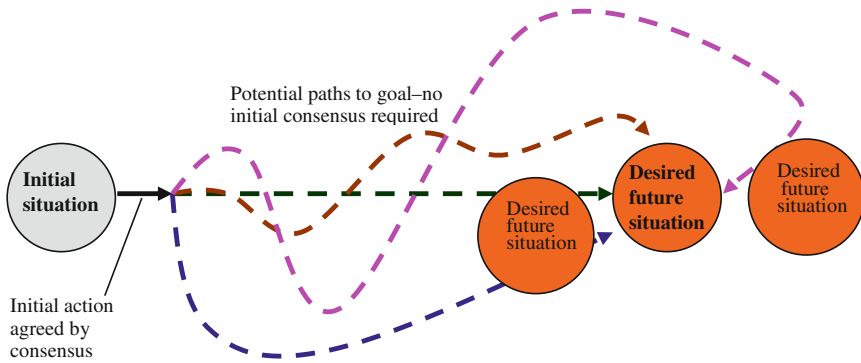
Fig. 3.2 Linear and nonlinear pathways to goal (adopted from Bakewell and Garbutt 2005)

The Necessity of Explicit Assumptions

Assumptions, potentially one of the most important aspects of the log frame, are also perhaps the most misused. Assumptions are specified in the fourth column of the log frame. The aim of specifying the assumptions is to identify the external factors whose presence or absence will (or are likely to) affect the success of the project. It will have become apparent during the analysis phase that the project alone cannot achieve all objectives identified in the log frame. These objectives will affect the project's implementation and long-term sustainability, but they lie outside its control. These conditions must be met, however, if the project is to succeed (European Communities 1999). Assumptions in the log frame are stated in terms of the desired situation (Jackson 2000). The main condition in assumptions is that they are outside the control of project managers; they should be outside of the project's intervention, yet critical to the project's success (European Communities 1999).

Preconditions differ from assumptions in that they must be met before a project can commence. For example, without the implementation of certain policy measures by a key partner, the project rationale may be undermined. The likelihood of these conditions being met should be estimated as part of assessing the riskiness of the project. Some preconditions and assumptions will be more critical to project success than others. These characteristics are represented in Fig. 3.3 an algorithm for assumptions.

It is important for each program stakeholder to write out his/her assumptions as an initial step in the process, then discuss them as a group. This not only increases clarity about stakeholders' views of how the process will work, but it helps them individually and as a group to identify major and minor assumptions, to judge the likelihood of the assumptions being sustained or invalidated in practice, and identify the assumptions as more and less critical to achieving the desired result(s). It may be necessary to make this an iterative process to help identify group consensus.
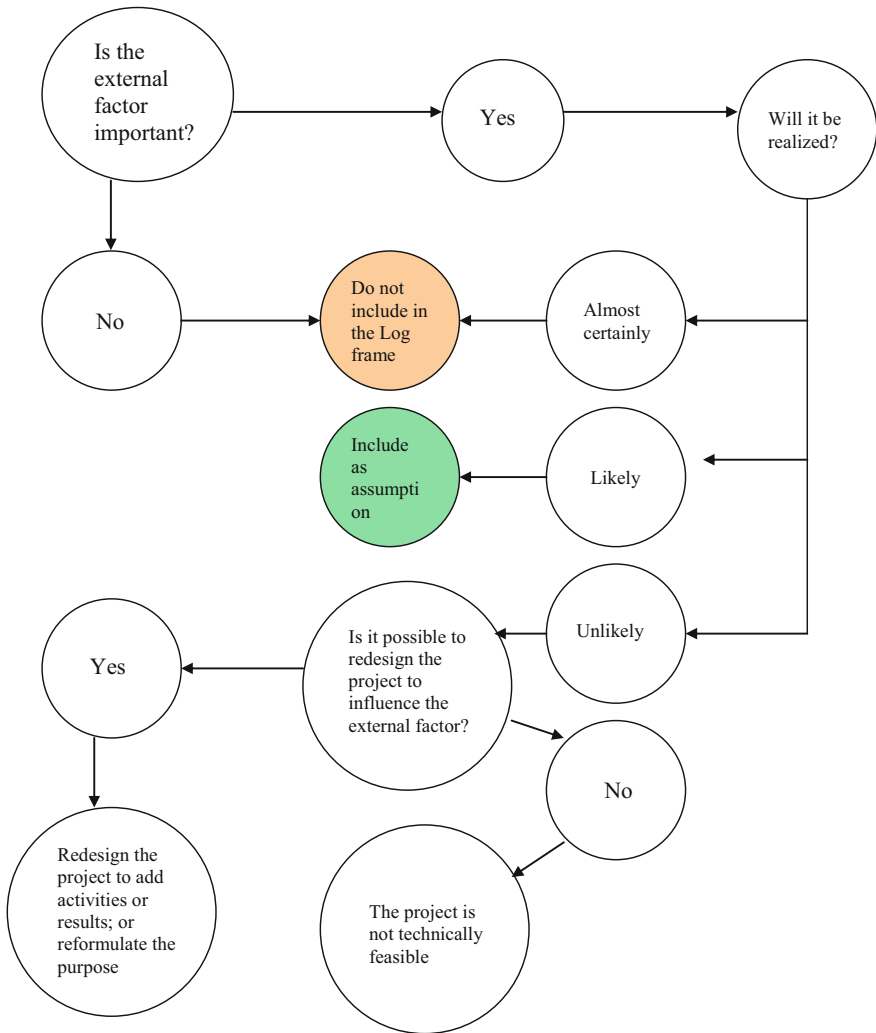
Fig. 3.3 An algorithm for assumptions (adopted from Jackson 2000)

Additionally, this process should make it easier to evaluate performance of such interventions. But depending on the rigor of the initial process, this may be more rhetoric than reality. Most of the assumptions we find in project log frames are not really helpful. They are mostly such truisms as, "We hope the project will achieve its purpose if the donor agrees to fund it;" or "There will not be outbreaks of civil wars," or "The community will be willing to participate." The table below provides examples of ill-designed assumptions gone awry.

It is true that if there is an outbreak of a civil war, an intervention aimed at increasing farmers' incomes by providing with them seeds is unlikely to succeed.
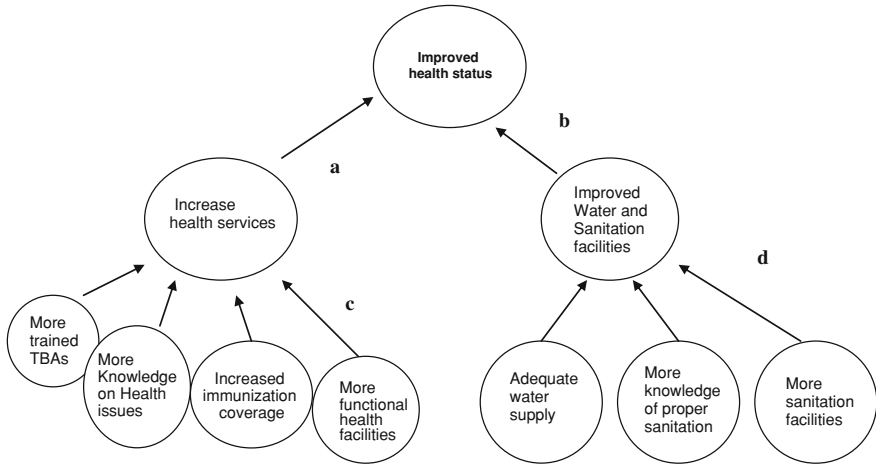
Fig. 3.4 Results chain of a health project

Linkage area	Assumptions stated	Some additional questions you would ask	Some observations
A&B	Depending on how "improved health status" is defined, increasing access to health services and water and sanitation facilities would improve the health status of community project participants.	Why do we believe that in this community, health depends mostly on availability of facilities?	This plan is silent about other aspects that influence peoples health, especially those related to the individual's responsibility for wellbeing
C	Trained Traditional Birth Attendants remain in the communities where they are posted Peace prevails communities are willing to be involved	Why do we think that training TBAs, increasing health awareness, immunization coverage and the number of functional health facilities are the most appropriate ways of raisings health services delivery? What else might hinder the effectiveness of these approaches and how would these limitations be addressed?	This plan is silent about quality of health services and how this could affect demand. Could it be an issue? The service oriented approach is more aligned with curative than preventive health; does the narrative clarify this choice?
D		This linkage shows that a combination of 'software' (education for behavioral change) and 'hard ware' (facilities) is needed for better sanitation. Not sure if the narrative recognizes that there could be barriers to behavior change and how this would be resolved among the more as well as less literate members of the community.	

It is worthwhile, however, to note the assumption that farmers who have more seeds will have more income. Depending on the course of the civil war, the assumption may hold true in general, for a major portion of the country, or only in scattered areas. Depending on the farmers' skill with a new crop within any given area, incomes may increase, decrease, or remain stable. This can be a starting point for all the assumptions that actually have to be true in order to achieve the desired outcome. For example, the farmers have to have safe storage for the seed until planting time; they need to know how to plant these particular seeds properly and how to care for the growing crops. If the crop is a new one, the farmer(s) and their family(ies) must have some idea about what its uses are. The assumptions need to be spelled out. If an assumption is not explicated, it cannot be explored for consequences, good, or ill.

Figure 3.4 shows an example of a project that hopes to improve the health status of people in a community by increasing health services delivered, which will be achieved through training more providers and developing more facilities, among other activities. In this project's log frame, the key assumptions listed were "No disease outbreaks" and "Support of government's health ministry." We should ask, "Why have providing health services and sanitation facilities been identified as the best ways to improve health in this community? Why choose these approaches over others?" Assumptions should not be wishes or hopes; they should explore in realistic ways what conditions are necessary and sufficient for success in meeting the project's goals and what changes—for better or worse—have some chance of occurring. The current assumptions column does not help us explicate the assumptions, let alone explore them.

What Else Might Hinder the Effectiveness of These Approaches (Or Support Their Success), and What Have We Thought About the Likelihood of Those Factors?

Figure 3.5 shows a project that seeks to improve household food security by increasing household food availability, as well as reducing post-harvest crop waste. Why do we believe that the food security question in this community is about household food availability and post-harvest waste? What else (other than the stated assumptions) might hinder the effectiveness of this approach and what have we thought around those factors. Why do we believe that the best way to improve household food availability in this community is by increasing acreage, incomes, and access to markets? Why do we choose these approaches over others? Is this rationale clear in the narrative? How do these pieces work together to bring about the intended change? To extend the analysis, does the project plan to measure village by village or province-wide? What differences could this make in the assumptions?
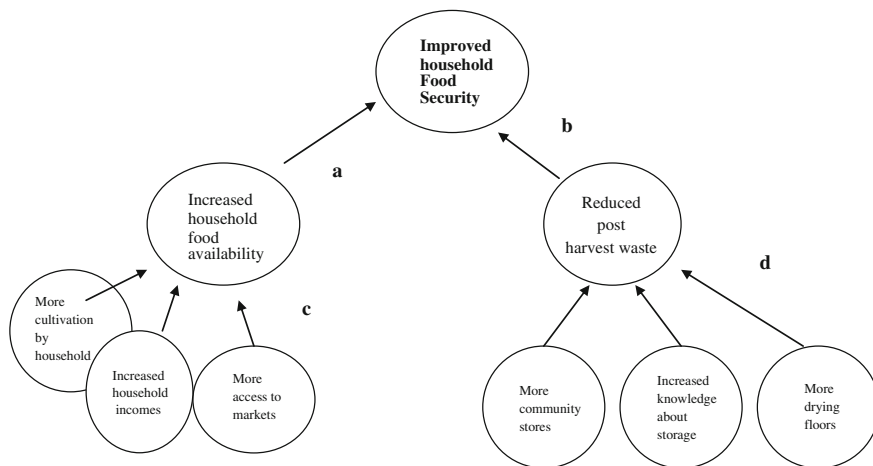
Fig. 3.5 Results chain of a food security project

Linkage area	Assumptions stated	Some additional questions you would ask (check consistence between narrative and log frame)	Some observations
A&B	Communities are wiling to adopt new technologies Communities will make material contributions	Why do we believe that the food security question in this community is about household food availability and post harvest waste? What else (other than the stated assumptions) might hinder the effectiveness of this approaches and what have we thought around those factors)	Increasing household incomes could flow from increased agricultural production and access to markets
C	As above	Why do we believe that the best way to improve household food availability in this community is by increasing acreage, incomes, and access to markets? (Why do we choose these approached over others? Is this rational clear in the narrative? How do these three pieces work together to bring about increase d food availability?)	Land tenure might be an issue here
D	As above	This linkage shows that a combination of 'software' (education for behavioral change) and 'hard ware' (facilities) is needed for better sanitation. Not sure if the narrative recognizes that there could be barriers to behavior change and how this would be resolved among the more as well as less literate members of the community.	

Gasper (2000) presents some interesting arguments about the superficial treatment of assumptions in log frames. First, he argues that it is not the problem of the log frame itself as it is the problem of the users (Gasper 2000, p. 17):

> LFA then at least shows up casualness about assumptions and prepares the way for more serious treatment.

But he also argues that there is a problem with the LFA:

> 'Yet it (the LFA) literally marginalizes assumptions analysis, locating it in the final column of the matrix. The fourth column is the second one to be defined when constructing a logical framework without noting the incongruity. The bulk of users who came to the assumptions column fourth tended to be tired and on too different a mental track, after an arduous traverse of indicators".

Another difficulty has been suggested in the way the LFA defines assumptions. They are considered external factors beyond the control of the project and outside the project intervention. Knowing such assumptions is critical, but may not help evaluators in explaining program performance. In the next section, we review another evaluation approach that takes a different perspective on explicating assumptions. This approach, the Theory of Change approach, is spawning a lot of interest, even some prominence, in the evaluation community.

Another weakness in the use of log frames is the inclusion of what Powell (2011) calls "redundant results" and Bury (2011) terms a "category error". This arises when a subsidiary result or outcome is placed below a higher level (composite) result or outcome, to which it should contribute, and yet in fact it is simply part of its definition. For example, a result such as increased awareness of HIV among schoolchildren and their parents might have below it the results "increased awareness of HIV among schoolchildren and increased awareness of HIV among parents." These two subsidiary results are redundant and should be removed. If they are the same thing as the composite result, they cannot cause it. Bury (2011) suggests that a good results framework ensures that the lower level results are (a) all necessary to achieve the higher level result, (b) sufficient to achieve the higher level result, and (c) plausibly cause achievement of the higher level result, rather than merely being categories thereof. Anderson (2004) talks of domino effects in a results framework. Domino results are those that will emanate from a result without having to do anything additional. For example, a result above parents' awareness of HIV may suggest that a result leads to reduction in risky behavior," which is not part of the definition of 'increasing parents' awareness of HIV."

Powell (2011) identifies a number of scenarios where subsidiary results are not logically independent of composite results:

(a) When they exactly overlap: If subsidiary results overlap exactly with the composite result, it is better to do away with the redundant lower layer and specify those results in the definition of the composite result.
(b) A composite result can include content that is not covered in the subsidiary results. If the result "increased parents' and children's awareness of HIV" were changed to "increased awareness of HIV among educational stakeholders",

the awareness levels of parents and children become *part* of the outcome, but the result also refers to teachers, principals, education authorities, etc. Now, do we assume that the subsidiary results (or any other results elsewhere in the framework) can make a *causal contribution* to this additional outcome among parents? If no, then there may be a component of this result that will have no programming to change and which should just be deleted along with the entirely redundant subsidiary level. If yes, because it is believed that the children and their parents will influence other stakeholders, then it should be clarified that there is an assumption that parents' and children's awareness will be multiplied by passing that awareness along to other stakeholders.

(c) The subsidiary results can include content that is not covered in any of the composite results. So if the result "Increased parent and children's awareness of HIV was changed to "increased awareness of Sexually Transmitted Infections among school children", parents' awareness of HIV is still part of the definition, but it also includes components that are not covered. It may be assumed that this additional content in the subsidiary results can make a *causal contribution* to the composite (or to any other results in the framework). If no, then there is content in the subsidiary results that does not lead to higher level change and which should probably just be deleted (and then the subsidiary level is entirely redundant and we can delete that too). If yes, perhaps because the assumption is deemed valid that children's knowledge of other STDs will influence or reinforce their or other stakeholders' awareness of HIV, again the basis of this expectation should be clarified and logically presented.

The Theory of Change Approach

'Theories of Change' (ToC) has proved a popular approach for both evaluators and commissioners of evaluations for complex social policy programs (e.g., Coote et al. 2004; Judge 2000; Mason and Barnes 2007; Jacobs et al. 2002; Weiss 2000). This approach has provided program developers with useful tools for working through designs of complex change processes, by helping them to clarify and reflect on how selected programming approaches and strategies can or should elicit desired results.

Much work on popularizing the ToC Approach (ToCA) has been done by The Aspen Institute's' Roundtable on Comprehensive Community Initiatives (RTCCI) and the Kellogg Foundation's work on community-building and improvement. These two organizations have approached the ToCA in similar ways.

According to RTCCI, a theory of change explains how a group of early and intermediate accomplishments sets the stage for producing long-range results. A more complete theory of change articulates the assumptions about the process through which change will occur and specifies the ways in which all of the required early and intermediate outcomes related to achieving the desired long-term change will be brought about and documented as they occur (Anderson 2004, p. 3). RTCCI's approach to theory of change requires stakeholders to be precise
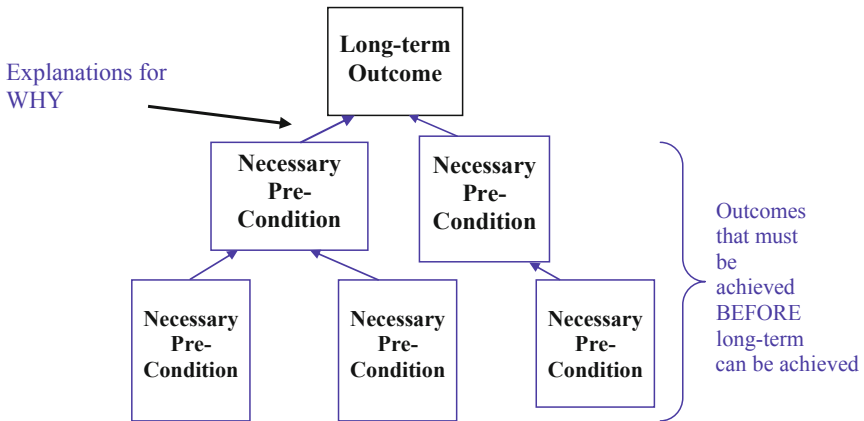
Fig. 3.6 Pathways of change (adapted from Clark and Anderson 2004)

and clear about the type of changes they want to achieve. Participants in theory of change workshops are asked to predict exactly who or what is going to change, over what period of time, and by how much, at every step in an often complex process. "We ask them to specify how and why they expect change to happen in a particular way. We also ask how they are going to bring their resources to bear on creating early and intermediate changes that add up to their ultimate goal. Simple questions, in theory…but difficult to answer in practice" (Anderson 2004).

RTCCI's ToC has four components:

(1) A pathway of change that illustrates the relationships among a variety of outcomes that are each considered preconditions of the long-term goal.
(2) Indicators that are defined to be specific enough to measure results—success or failure or some intermediate state.
(3) Interventions that are used to bring about each of the preconditions on the pathway, and at each step of the pathway.
(4) Assumptions that explain why the whole theory makes sense.

Pathway of Change

A pathway of change is a map that illustrates the relationship(s) between actions and outcomes and also shows how outcomes are related to each other over the lifespan of the initiative (See Fig. 3.6). The long-term goal of the initiative appears at the top of the map, and the outcomes that must be produced in order to achieve the goal are arranged in sequence with the earliest at the bottom and the final outcomes at the top of the list. This map is then read from the bottom to the top, suggesting that the earliest outcomes (at the bottom) are needed to arrive at the next level, and outcomes at the middle level are needed to reach the top (Anderson 2004).

Outcome and Precondition

For RTCCI, everything in the pathway of change is a precondition to the long-term goal. That is, the outcomes on the path are all required to reach the goal. Without each of them in place, we assume the goal cannot be attained. It is intended that this logic helps to weed out extra outcomes that may be nice but unnecessary to achieve the intended goal. It is not necessarily different from the results chain conceptualized in a log frame. What is different is that the ToCA asks whether these outcomes or preconditions are necessary and sufficient to reach the long-term goal.

Indicators

Indicators, another component that ToCA and LFA have in common, should help in recognizing success within the pathway of change. The difference is that with in ToCA, the indicator (for both the goal and outcomes), might include more detail that "operationalizes its definition" (Anderson 2004). The indicators are operationalized by these questions: (a) who or what is the target population for change? (b) How much change has to occur in this indicator for us to claim to have successfully reached the outcome? (c) How long will it take to bring about the necessary change in this indicator in the target population? These questions are essential in elaborating indicators as a component of the program's ToC.

Intervention

This component describes the types of actions that would be required to bring about each of the preconditions on the pathway of change. This may well be similar to inputs and outputs on the log frame. The difference might be that in the ToCA, an intervention might be as simple as a single activity or as complex as an entire program. Each outcome in the pathway of change is linked to a specific intervention, revealing the often complex web of activity that is required to bring about the desired long-term community change (Anderson 2004).

Assumptions

This is another component that ToCA and LFA have in common, though the strategies conceptualize their roles differently. In the ToCA, assumptions explain both the connections between the preconditions for long-term change occurring in the early and intermediate stages and the expectations about how and why proposed interventions will bring them about (Anderson 2004). In LFA, the

Table 3.2 Comparing log frame and theory of change

Log frame	Theory of change
Logic models graphically illustrate program components, and creating one helps stakeholders clearly identify outcomes, inputs, and activities	Theories of Change link outcomes and activities to explain HOW and WHY the desired change is expected to come about
Logic models usually start with a program and illustrate its components	Theories of change may start with a program, but are best when starting with a goal, before deciding what programmatic approaches are needed
Logic models require identifying program components, so you can see at a glance if outcomes are out of sync with inputs and activities, but they does not show WHY activities are expected to produce outcomes	Theories of change also require justifications at each step—you have to articulate the hypothesis about why something will cause something else (it is a causal model, remember!)
Logic Models does not always identify indicators (evidence to measure whether outcomes are met or not) (show that they identify indicators differently)	Theories of Change require identifying indicators that will appropriately monitor desired changes/outcomes.
Logic Models are great when you need to:	Theories of Change are best when you need to:
Show someone something they can understand at a glance	Design a complex initiative and want to have a rigorous plan for success
Demonstrate you have identified the basic inputs, outputs, and outcomes for your work	Evaluate appropriate outcomes at the right time and the right sequence
Summarize a complex theory into basic categories	Explain why an initiative worked or did not work, and what exactly went wrong

Adapted from Clark and Anderson (2004)

assumptions are always unstated, despite the assumptions column. In LFA, assumptions are those factors that are external to the program which could affect the success of the program. These two framings of the assumptions concept are both important. It seems, though, that those developing these development programs have been more schooled to the concept that assumptions are external, and that alone does not help evaluators understand programs performance. ToCA gives more attention to this long-overlooked issue of assumptions that affect, or could affect, the internal structure and operations of the program. These assumptions may be supported by research, or "best practices," that can strengthen the case for the plausibility of theory and the likelihood that stated goals will be accomplished.

Three main types of assumptions underlie a theory of change:

1. Assumptions about why each precondition is necessary to achieve the result in the pathway of change and why the set of preconditions is sufficient to bring about the long-term outcome.
2. Assumptions from social science theory that connect program activities to specific outcomes for specific populations or communities. This may include findings from "best practice" research as well as evidence from academic (or basic) research.

3. External assumptions about the context/environment in which the program or intervention will be implemented.

When a theory of change is built around the wrong assumptions about the project's context or environment, even the most elaborate pathway of change can fall apart once the incorrect assumptions manifest themselves—not to mention that a plan based on faulty assumptions is not likely to bring about the desired goal. Therefore, it is important to have the stakeholder group think critically about what they consider to be true about the operating environment as well as the links in the theory of change before they sign off on their theory (Anderson 2004).

RTCCI offers some probing questions that participants may use to elaborate and check assumptions in the pathways of change:

- When you look at the total picture, do you believe that the theory makes sense?
- Do the preconditions make sense as the logical steps toward the long-term outcome?
- How will we be able to bring about the outcomes at the levels we have predicted?
- Is there anything going on in the real world that may make it difficult to get this theory off the ground the way we have planned it?
- Is this theory of change PLAUSIBLE? Have we created a compelling story about the pathway of change that would lead to the long-term goal in this community?
- Is this theory of change FEASIBLE? Do we have the capacities and resources to implement the strategies that would be required to produce the outcomes in the pathway of change?
- Is this theory TESTABLE? Have we specified how success will be measured clearly enough that we can recognize progress toward our goal when we see it? Have we defined indicators for each outcome in clear terms that a researcher or evaluator can use to produce a research plan?

Theory of Change Versus Log Frame

There has been much discussion (and confusion) about differences between log frame and ToC. The table above summarizes the main differences (See Table 3.2).

An illustration by Anderson and Gina (2004) helps to further illuminate the difference between a ToC and log frame. How often do students need to attend an after-school program in order for their reading tests scores to improve? A Logic Model would tell you that the after—school program is an activity and improved reading scores are an outcome. It might tell you that attendance at the after—school program is an intermediate outcome. But it would not tell you that "Students need to attend after-school programs at least three days per week for a minimum of 60 days, and the curriculum must focus on love of reading and literacy in order for test scores to rise."

Participatory Impact Pathways Analysis

Participatory impact pathways analysis (PIPA) is an approach to planning, monitoring, and evaluation developed for use with complex research-for-development projects (See Douthwaite et al. 2003, 2007, 2008; Renger and Titcomb 2002). PIPA begins with a participatory workshop where stakeholders make explicit their assumptions about how their project will make an impact, and then produce an "outcomes logic model", and an "impact logic model."

These two logic models provide an ex-ante framework of impact predictions that can also be used in priority setting and ex-post impact assessment. PIPA engages stakeholders in a structured participatory process, promoting learning and providing a framework for "action research" on processes of change (Alvarez et al. 2010).

The PIPA approach enables project staff and stakeholders to describe jointly the project's theories of action, develop logic models, and use them for project planning and evaluation. The term impact pathways is synonymous with the terms "theories of action" and "program theory". Similar to ToCA and LFA, PIPA encourages stakeholders to think through how project activity outcomes can lead to intermediate and long-term changes. With the PIPA approach, a project's impact pathways are described in two ways:

1. A problem or objectives tree which describes the linear logic, showing a chain of results or sequence of changes expected from a project's outputs, ultimately leading to the goal.
2. Network maps, showing how project stakeholders work together, influence each other, and influence the general environment; all of which influence attainment of project results.

The approach utilizes a workshop process that is designed to help participants raise, describe, and discuss their hypotheses for how project activities and outputs could eventually contribute to desired goals. Within the workshop, stakeholders are facilitated to develop the two perspectives shown in Fig. 3.7, and then integrate them through developing an outcomes logic model that describes the project strategies, outputs, and outcomes necessary to achieve the project vision. The outcomes logic model links the outcomes to the actor or group of actors that will bring them about.

The description of these hypotheses is a description of the project's impact pathways. Participants develop these pathways by:

1. Clarifying, reaching mutual understanding, and communicating their project's intervention logic and its potential for achieving impact;
2. Understanding other projects working in the same program, and identifying areas for collaboration;
3. Generating a feeling of common purpose and better programmatic integration (when more than one project is represented in the workshop);
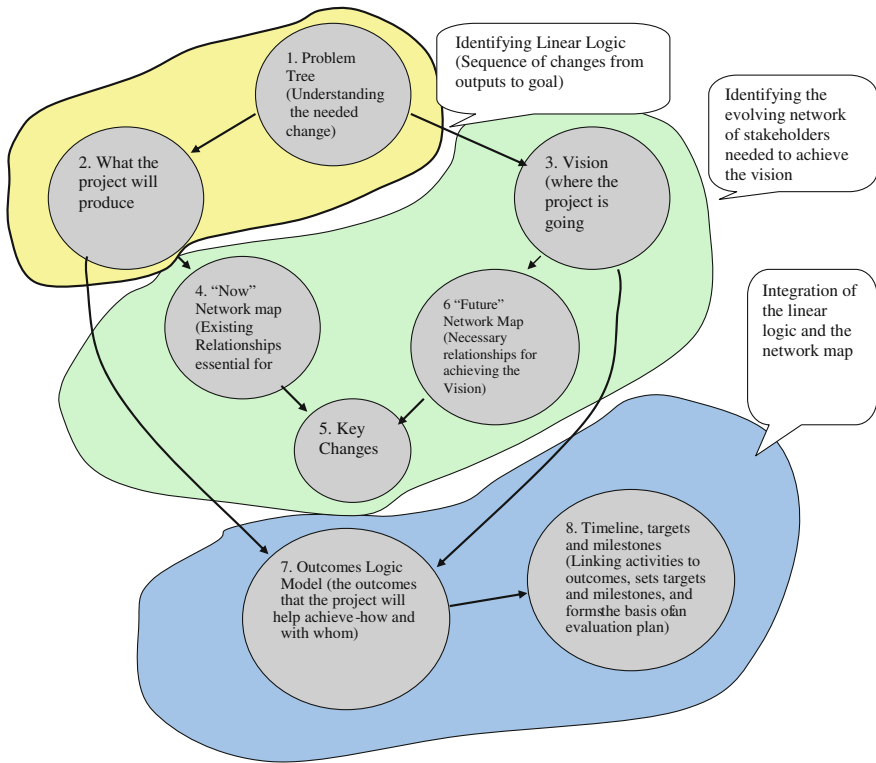
Fig. 3.7 The PIPA workshop (adapted from Alvarez et al. 2010)

4. Developing a narrative describing the project's intervention logic and possible future impacts (thus a form of ex-ante impact assessment);
5. Developing a framework for subsequent monitoring and evaluation.

There are six main steps in the process:

Step 1: Participants develop a problem tree which links the problems that the project is directly addressing with the social, environmental, and economic conditions that it wishes to improve.

Step 2: The problem tree is turned into an outcome tree that represents the situation after problems have been addressed.

Step 3: The outcome tree is converted into an outcomes logic model that describes changes in knowledge and attitudes and skills and practices.

Step 4: Strategies to bring about these changes are devised and articulated.

Step 5: A monitoring and evaluation plan and a schedule of activities are developed.

Step 6: The plan is executed.

The changes listed in the outcomes logic model are ones that are possible, at a stretch, to achieve within the timeframe of the project. They generally describe the

expected direct benefits of the project and can be evaluated through the monitoring and evaluation system described above.

Comparing PIPA, LFA, and ToC Approaches

All three approaches—PIPA, LFA, and ToC—emphasize participatory processes in which stakeholders identify problems and devise strategies for achieving outcomes and outlining pathways by which these outcomes would be achieved. However, completing formats of a traditional logical framework is difficult to do in groups and is thus often dominated by one or two people (Douthwaite et al. 2007). Groups are less constrained when they use flow diagrams used in PIPA and ToC processes to map pathways of change; the skills of a facilitator not withstanding. In the PIPA workshop, participants only attempt to create a logic model once the underlying impact pathways have been discussed and agreed.

The impact logic model is a flowchart that shows the chains of outcomes that link outputs to eventual developmental impacts. The other element that makes PIPA useful is the use of network maps to complement logic models in explaining impact pathways. A net work map is a flow chart that shows the chains of outcomes that link outputs to eventual developmental impacts. The network maps give additional detail to the causative theory, given the principle that the interactions among stakeholders are key to the scale-up and scale out of intervention effects (Douthwaite et al. 2007). Network maps are used to explore how stakeholders are linked to and influence each other, and how the project aims to change the existing networks (Douthwaite et al. 2007).

All three approaches—LFA, ToC, and PIPA—provide opportunities for unearthing stakeholder assumptions. The LFA approach tends to focus more on external assumptions—the factors beyond the control of project implementers that influence the success of the project. While those are important assumptions to consider, the assumptions related to how program outputs will be transformed into desired long-term changes are often not unearthed, and these are more dealt with by the ToC and PIPA approaches. The reflection on both the results and relationships aspects of planned change with in PIPA provides many opportunities for unearthing stakeholder assumptions.

It is also important to note that creativity and facilitation skills account for much of what any of these three approaches can provide in designing complex development programs; sometimes it may necessitate a mixture of more than one design approach depending on the complexity of program's internal and external environment. Just as the choice of a program design approach needs to respond to contextual complexity; the choice of program evaluation methodology needs to be appropriate to a program's complexity. The next chapter discusses the need to navigate the evaluation methodology toolbox guided by the "appropriateness" principle.

References

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administration Science Quarterly, 45*(1), 425–455.

Abonyi, G., & Howard, N. (1980). A Boolean approach to interactive program planning. *Management Science, 26*(7), 719–735.

Alvarez, S., Douthwaite, B., Thiele, G., Mackay, R., Co′rdoba, D., & Tehelen, K. (2010). Participatory impact pathways analysis: a practical method for project planning and evaluation. *Development in Practice, 20*(8) 946–958.

Ambrose, K. & Roduner, D. (2009). *A conceptual fusion of the logical framework approach logical framework approach and outcome mapping.* OM Ideas Paper No. 1.

Anderson, A. (2004). *Theory of change as a tool for strategic planning: a report on early experiences*. Aspen Institute: Washington, DC.

Bury, B. (2011). Response to Steve Powell on illogical frameworks, composite results and logframe bloat. Retrieved on December 12 2011, from socialdatablog.com.

Chandrasekhar, A. G., Kinnan, C. & Larreguy, H. (2011). *Informal insurance, social networks, and savings access: Evidence from a lab experiment in the field.* MIT Working Paper.

Clark, H., & Anderson, A. (2004, November). *Theories of change and logic models: Telling them apart*. Presentation at American evaluation association annual conference. Atlanta, Georgia.

Coote, A., Allen, J., & Woodhead, D. (2004). *Finding out what works*. London: Kings Fund.

Douthwaite, B., Schulz, S., Olanrewaju, A. S., & Ellis-Jones, J. (2003). Impact pathway evaluation: an approach for achieving and attributing impact in complex systems. *Agricultural Systems, 78*, 243–265.

Douthwaite, B., Alvarez, S., Cook, S., Davies, R., George, P., Howell, J., et al. (2007). Participatory impact pathways analysis: a practical application of program theory in research-for-development. *The Canadian Journal of Program Evaluation, 22*(2), 127–159.

Douthwaite, B., Alvarez, S., Thiele, G., & Mackay, R. (2008). *Participatory impact pathways analysis: A practical method for project planning and evaluation* ILAC Brief May 17 2008 The Institutional Learning and Change (ILAC) Initiative (www.cgiar-ilac.org).

Hummelbrunner, R. (2010) Beyond logframe: Critique, variations and alternatives in Nobuko Fujita, beyond logframe; *Using systems concepts in evaluation*, Foundation for advanced studies on international development, Tokyo, pp. 1–34.

Bakewell, O., & Garbutt, A. (2005). *The use and abuse of the logical framework a review of international development NGOs' experiences*. Stockholm: SIDA.

Davies, R. (2004). Scale. *Complexity and the representation of theories of change, evaluation, 10*(1), 101–121.

Davies, R. (2005). *Moving from logical to network frameworks: A modular matrix approach to representing and evaluating complex programs, in Creative Commons Attribution-NonCommercial-ShareAlike site* available online at: http://66.102.9.104/search?q=cache:cs98bp1tuXgJ:www.mande.co.uk/docs/MMA.doc+network+software+Visualyzer&hl=en&ct=clnk&cd=2&gl=uk

Deprez, S., Van Ongevalle, J. (2006a). Building reflection and learning into educational programmes, introducing outcome mapping—the case of St2eep. Conference Proceedings, International Conference on Strengthening Innovation and Quality in Education, Leuven, Belgium.

Deprez, S., Van Ongevalle, J. (2006b) *Outcome mapping: Learning the way forward—an alternative way to plan, monitor and evaluate programmes. Conference Proceedings, Environmental Education Association of Southern Africa.* Harare, Zimbabwe.

Earl, S., Carden, F., & Smutylo, T. (2001). *Outcome mapping: Building learning and reflection into development programs*. Ottawa, ON: International Development Research Centre.

European Communities (EC) (1999). *Project cycle management training handbook*. Brussels: Author.

Gasper, D. (1997). *'Logical frameworks': A critical assessment*. Working Paper no. 278. The Hague: Institute of Social Studies.

Freeman, L. (2006). *The development of social network analysis*. Vancouver: Empirical Press.

Gasper, D. (2000). Evaluating the 'logical framework approach' towards learning-oriented development Evaluation. *Public Administration and Development., 20*(1), 17–28.

Jackson, B. (2000). *Project designing projects and project evaluations using the logical framework approach*. Retrieved October 12 2011, from www.Management/logicalframeworkapproach.htm.

Judge, K. (2000). Testing evaluation to the limits: The case of english health action zones'. *Journal of Health Services Research and Policy, 5*(1), 1–3.

Hogan, B., Carrasco, J., & Wellman, B. (2007). Visualizing personal networks: Working with participant-aided sociograms. *Field Methods, 19*(2), 116–144.

Jacobs, B., Mulroy, S., & Sime, C. (2002). Theories of change and community involvement in North Staffordshire health action zone. In L. Bauld & K. Judge (Eds.), *Learning from health action zones*. Chichester: Aeneas Press.

Johnson, M. A., Casillas, W., Brown, J. U., & Trochim, W. (2011, November). *Using systems thinking in evaluation capacity building: The systems evaluation protocol.* Paper Presented American evaluation association annual conference, Anaheim CA.

Kibel, B. (2000). *Outcome engineering toolbox: User manual.* Retrieved June 15 2011, from Pacific Institute for Research and Evaluation Web site: http://www.pire.org/resultmapping/homepage.htm.

Leischow, S. J., Best, A., Trochim, W. M., Clark, P. I., Gallagher, R. S., Marcus, S., et al. (2008). Systems thinking to improve the public's health. *American Journal of Preventive Medicine, 35*(2), 196–203.

Mason, P., & Barnes, M. (2007). Constructing theories of change: Methods and sources. *Evaluation, 13*(2), 151–170.

Moody, J., & White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review, 68*(1), 103–127.

Owen-Smith, J., & Powell, W. W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston Biotechnology community. *Organization Science, 15*(1), 5–21.

Powell, S. (2011). *Illogical frameworks, composite results and log frame bloat.* Retrieved December 12 2011, from socialdatablog.com.

Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *American Journal of Evaluation, 23*, 493–503.

Richard, R. F. (2009, November). *The logic model and systems thinking: Can they co-exist?* Paper presented at the American Evaluation Association Conference. Orlando FL.

Rockwell, S. K., & Bennett, C. F. (2000). *Targeting outcomes of programs*. Retrieved July, 2011 from University of Nebraska, Institute of Agriculture and Natural Resources. http://deal.unl.edu/TOP/

Rowlands, J. (2003) Beyond the comfort zone: some issues, questions, and challenges in thinking about development approaches and methods. In D. Eade (ed.), *Development methods and approaches: Critical reflections. A development in practice reader* (pp. 1–20). Oxfam GB, London.

Trochim, W. M., Cabrera, D. A., Milstein, B., Gallagher, R. S., & Leischow, S. J. (2006). Practical challenges of systems thinking and modeling in public health. *American Journal of Public Health, 96*, 538–546.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Williams, B., & Iman, I. (2008). (eds.). *Systems concepts in evaluation: An expert anthology*. Point Reye, CA: Edge Press of Inverness.

William, B. (2011). Comments on a blog. http://www.cognitive-edge.com

Weiss, C. (2000). Which links in which theories shall we evaluate? *New Directions for Evaluation, 87*(Fall), 35–45.

World Bank (2009). Interactive community planning-upgrading Urban communities. Retrieved September 23 2011, from http://web.mit.edu/urbanupgrading/upgrading/issues-tools/tools/ZOPP.html.

Chapter 4
Evaluating Complex Development Programs

Abstract Evaluators express preferences for certain methods over others. This chapter highlights the debate and assumptions underlying these preferences. There are debates on which methods represent the "gold standard" for evaluation. The key point discussed in this chapter is that the "gold standard" for evaluation methodology is *Appropriateness* to evaluation questions and the contextual realities of complex programs.

> *In God We Trust—All Others Must Have Credible Evidence*
> Donaldson et al. 2009, p. 2

> *Neither the quantitative hook set of for the big fish nor the qualitative net scaled for the little fish adequately captures life in the most seas. We need a paradigm to help us become scuba divers*
> Dattu 1994, pp. 61–70

> *It has been said that economics is a box of tools. But we shall have to resist the temptation of the law of the hammer, according to which a boy, given a hammer, finds everything worth pounding, not only nails but also Ming vases. We shall have to look, in the well known metaphor, where the [lost]key was dropped rather than where the light happens to be. We shall have to learn not only how to spell 'banana' but also when to stop. The professionals, whom a friend of mine calls 'quantoids' and who are enamored of their techniques, sometimes forget that if something is not worth doing, it is not worth doing well.*
> Streeten 2002, p. 110 quoted in Debrah Eade 2003, p. ix

> *With in any given approach, how do we choose the methods we use? And what factors influence how we apply them? These choices are not always clear or conscious. If we are really honest about it, many of us probably often choose methods*

51

because they are familiar and draw on skills that we feel confident about; or because we perceive that the funders require them……realistically, most 'practitioners' operate within constraints of time and resources that affect the choices we make; we also operate within political contexts that shape our choices

Rowlands 2003, p. 3.

Introduction

In the face of complex evaluands, evaluators encounter a key question: Which methods to use? It is not uncommon that evaluators fill their tool box preferentially for reasons different from the nature of problem at hand. Evaluators have not disagreed about the complexity of development programs as they have disagreed about which approaches to use in evaluating such programs. The debate on approaches (paradigms and methods) has been very engaging, and it does not seem that this debate is about to come to its conclusion. It is worth noting that evaluators have come a long way with this debate, and there is much agreement on the need for mixed methods (Hughes and Hutchings 2011) or methodological eclecticism (Roberts 2002). However, the methodological debate persists, or some evaluators just feel this way. For example, the most recent volume of New Directions for Evaluation (A journal of American Evaluation Association)-spring 2012) has a focus on this debate. As Julnes notes:

> The impetus for this work is the recognition that there is little consensus in the evaluation community regarding a critical aspect of what we do—what methods should be used to help judge the value of public programs and policies? (Julnes 2012a, p. 1).

In the same volume, Julnes notes that "this lack of consensus on methods of valuing is becoming more problematic now that (a) evaluation is becoming more central to public-sector decision making and (b) the increasing pressure for evidence-based governance is pushing for more evidence-based, and hence systematic, policies on the methods of valuing appropriate for evaluation, often privileging specific approaches to assessing performance and economic impacts" (Julnes 2012b, p. 4).

A not-so-accurate description of this debate may look as follows: On one side of the isle in this debate is the "positivist", "science" "quantitative" and "expert" perspective (I refer to these as the dominantly quantitative perspective-DQuant). On the other end of the aisle are the "naturistic", "post-positivist" "constructivist", "Qualitative", "Art", and "stakeholder-focused" perspectives (I refer to these as the Dominantly Qualitative Perspective-DQual).

Although mixed methods approaches are increasingly advanced and popularized, many an evaluator might belong to either the DQuant or DQual camps. And it is interesting how we all get schooled to these camps. House (1984, p. 21) uses

this analogy: "Imagine Baptists teaching Catholic doctrine in their churches" to explain that evaluators can hold strong views for or against certain methods in a way comparable to how people may hold certain belief systems. Chelimsky (2012) connects methodology to evaluator values. Her argument is that evaluation is about valuing, a process that "is complex, diffuse, emotionally laden, and loaded with euphemisms" (p. 77). The question then is that of how to measure the value of a program. How is this methodological question linked to evaluator values? Chelimsky explains that "stronger methods allow us more confidence in the value judgments we make about a particular intervention, and weaker ones (which may be the only alternatives feasible in a particular design situation) force us to pepper those judgments with needed caveats". This shows that, like beliefs and values, an evaluator's orientation to an approach may not be easy to change. Perhaps, this may be the reason-or one of the reasons why this methodology debate persists.

The DQuant Camp

The DQuant perspective has always been referred to as "scientific method." Scientific method is rooted in Newtonian physics and it embraces "all the methods of observation, deduction, observation, generalization and the rest that had been found appropriate to the problems and issues preoccupying those subjects" (Toulmin 2001, p. 1). Thus, scientific method gives scholars a confidence in their established procedures for solving life's problems. Fay has put it this way: "We prefer our intellectualism expressed in the more sober tones and nuances of semantics." (Fay 1975, cited in Ball 1995, p. 266).

Science also assumes that the universe is a vast single system in which the basic rules are everywhere the same (Association for the Advancement of Science (AAS) 1990, p. 2). Thus, knowledge gained from studying one part of the universe is applicable to other parts. This is the principle of objective reality—that events or phenomena can be measured in some way (Ball 1995). For instance, the same principles of motion and gravitation that explain the motion of falling objects on the surface of the earth also explain the motion of the moon and the planets (AAS 1990).

Scientific study involves holding a number of other factors constant in order to examine quantitatively the relationship between one chosen input factor and a particular output (Wrigley 2004a, b). This principle is known as reductionism. Ayala (e.g., Ayala 1994) proposes three types of reductionism: methodological, epistemological and ontological.

1. Methodological reductionism implies that the best scientific strategy is to attempt to reduce explanations to the smallest possible entities.
2. Epistemological reduction implies that processes, properties, laws or theories found in higher levels of complexity can be derived entirely from those found in lower levels of complexity.

3. Ontological reductionism claims that higher-level, more complex entities are nothing but complex organizations of simpler entities, i.e., the whole is nothing but the sum of its parts.
4. Another two forms of reductionism have been suggested by Murphy: causal and reductive materialism (see Murphy et al. 2009).
5. Causal reductionism asserts that all causes are bottom-up; the characteristics and processes of the parts entirely determine those of the whole.
6. Reductive materialism implies that the entities at the lowest level are actually real, whereas higher level entities are only composite structures made of atoms.

An evaluation approach that is commonly considered to represent principles of the scientific method is the Randomized Control Trials (RCTs). The DQuant camp considers RCTs to be the "gold standard" for evaluating the effectiveness of social programs (Handa and Maluccio 2010; Dunn 1998; Desrosieres, 1998; Newman et al. 1994; Zhu 1999; Shadish et al. 2002; Shadish et al. 1991; Hughes and Hutchings 2011).

The DQual Camp

On the other hand, the DQuals argue that it is unrealistic to measure social change. They emphasize the need to use qualitative and participatory methods (Cronbach 1982; Guba and Lincoln 1989; Lay and Papadopoulos 2007). The DQual camp holds the belief that subjective reality-the meaning of events to the individuals concerned and the inherently inefficel or unmeasurable is what matters most (Dattu 1994). The paradigm is characterized by the emphasis on rich, comprehensive, in-depth understanding of what has happened as the meaning of events to those involved (Dattu 1994).

DQuals emphasize the role of art in program evaluation, which is about "profound understanding, about vicarious experience, about our shared and unshared histories, about hopes and fears and dreams and anxieties, about appreciating and valuing, about ourselves and others as cultures" (Geertz 1983 in (Lincoln 1991, p. 2)

What the DQuants Think of the DQuals

From the standpoint of the positivism or the quantitative paradigm, a number of implicit assumptions have been made about the qualitative paradigm (Lincoln 1991)

- Science means rigor and art means the oftentimes messy judgment and ineffable part. With in science, rigor is about method (the tradeoff is between 'big'

method for more rigor versus realities (such as cost and relevance) or 'small' method for less rigor still in face of realities.

- It has been believed that art is secondary to science. There is a belief entrenched in the scientific community that art is not subject to the conventions of scientific method.
- There is an assumption that science is about knowing or rationality and art is about feeling and emotionality is taken as valid on its face, yet many instances exist where science and art of various sorts have combined powerfully.

What the DQuals Think of the DQuants

Program planning and evaluation approaches that have roots in scientific experimentation thus tend toward reductionism. Fay (1975) in Ball (1995 p. 259) uses the term 'policy engineering' to refer to this kind of simplification—outlining the results chain, i.e., those that seek the most technically correct answer to political problems in view of available social scientific knowledge are the policy engineers (Ball 1995). Both Fay (in Ball 1995 p. 259) and Ball (1995) argue that this is advantageous as it eliminates certain biases-that Fay refers to as 'the sublimation of politics' and Ball argues that this way, policy is "depoliticized and thoroughly technicalized." According to Ball, "the debates and conflicts which link policies to values and morals are displaced by bland rationalist empiricism, and the best we can aspire to is to be integrated critics" (p. 259).

However, this approach has been criticized as being unsuited for complex programs.

The assumption that other factors can be held constant is unrealistic in the real world. With a rather cynical tone, Dupre (2001) uses the term *scientism* to refer to the DQuant perspective: the idea that any question that can be answered at all can be answered by science. *Scientism* presumes that the things and events in the universe occur in consistent patterns that are comprehensible through careful, systematic study. Scientists believe that through the use of the intellect and with the aid of instruments that extend the senses, people can discover patterns in all of nature. Dupre explains that this culture in which there is an "exaggerated" and often distorted conception of what science can be expected to do or explain persists today (Dupre 2001, p. 1). According to Ball (1995), "the epistemic assumptions of order, structure, function, cause and effect are variously mobilized to represent 'the social' and in doing so, work to exclude the mobile, complex and ad hoc messy and fleeting qualities of lived experience. We become locked into the simple but powerful and most productive assumption that 'the social' is susceptible to parsimonious and orderly totalizing conceptions" Ball (1995), p. 259.

Wrigley (2004a, p. 229) has used the term "ecological and social damage" to refer to the danger reductionism imposes on social programs. Arguing that mechanistic causality, the belief in one-to-one correspondences, and the limited

attention to environmental influences and effects when tracing causal relationships within a system are features of reductionism that specifically undermine complexity. According to Anderson (1999), it is not acceptable to compress complex behavior in order to comprehend it. Chaos theory demonstrates that simple laws can have complicated, unpredictable consequences; and complexity theory describes how complex causes can produce simple effects (Anderson 1999).

The complex nature of development programs makes it difficult to process all the relevant information for scientific experimentation (Murphy et al. 2009). "Complexity points to something which is just beyond our ability to understand and control, yet we presume it is densely packed, ordered and structured in some way that we fail to comprehend as yet" (Nowotny 2005, p. 15). According to Warren (2011), planning complex program with the lack of information is like myth making:

> The issue is this: 98 % of the relevant information we need to write a good plan is unknown at the time we write it. We more or less know where we want to go, but the map we draw to get there is based on heroic assumptions about the terrain ahead well beyond our vision. We will certainly have gathered as much information as possible to help us create this map—government statistics about the project area, analyzed results from our own community surveys and focus group discussions, documentation from meetings with community leaders, and so forth. We may even have multiple scenarios mapped out to manage multi-dimensional risks that give our map a sense of flexibility to likely changes in conditions. But even with all this, a constant stream of unforeseen events, new information, and unpredicted behavior will quickly unravel our confidence in the assumptions on which our plan is based Warren (2011, p. 1).

He argues that reality is so complex that our plans can never foresee all preconditions for program success:

> There is a saying, 'reality always wins.' Unexpectedly heavy rains will halve the planned number of community participants in workshops over the winter. A report showing the community's school to be lagging will suddenly shift their priority from the health sector to education. The death of a key community leader will drain the momentum out of a whole line of activities that she had championed. Election violence will create security risks that will keep staff out of the community for three weeks. A string of unannounced donor visits will throw off an entire month of work, leaving staff playing catch-up for the quarter. The head office will announce changes in the financial software and chart of accounts for the coming fiscal year that will require several members of the administration team to attend training workshops, and they will then fall behind on requisitions and payments, forcing delays in planned development activities. Two staff members will give birth and three will get sick in an oddly coordinated attack on productivity right in the middle of the annual planning season. The dollar will weaken unexpectedly and suck 5 % of the budget out from under foot. At times it will feel like Murphy's Law is in effect. Reality isn't polite, it doesn't wait for an invitation, and it barges in. Reality always wins (Warren 2011, p. 1).

But Toulmin (2001) argues that this confidence in scientific method and reductionism has waned today, especially in social sciences, under the influence of complexity. The growing discontent with the application of experimentalism is expressed in various ways. One of them, by Chen and Garabe (2011) has outlined some of the lessons learned from this approach, which he calls top-down:

- The conditions for designing and administering an evidence-based intervention often do not resemble or are irrelevant to real-world operations.
- Evidence-based interventions do not address practical or service issues highly relevant to stakeholders.
- Evidence-based interventions are difficult to implement in the real world.
- An evidence-based intervention does not necessarily imply the intervention would likely be real-world effective (pp. 95–98).

Hughes and Hutchings (2011) have argued that organizations operating in the international development sector need credible, reliable feedback on whether their interventions are making a meaningful difference, but they may not have time, resources, and expertise to conduct experimental studies such as randomized control trials. There is continued debate in the field of evaluation on the extent to which experiments should be applied in program evaluation, and what should be the gold standard. It appears that the 'gold' standard in evaluation has to do with what is the most appropriate methodology for the questions and context of the program. Complexity implies the best evaluation approach is one that takes into account the dynamic interplay of factors that determine program outcomes.

An Example of the Paradigm Fights

In their book *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Donaldson, Christie & Mark (Donaldson et al. 2009) have described an example of overt conflict between the two gangs, which was triggered in 2003 when the U.S. Department of Education's Institute of Education Sciences declared a "wholesale commitment" to privileging experimental and some types of quasi-experimental designs over other methods in applied research and evaluation unding competitions. An official response to this decision from the American evaluation association (AEA) stated that the implementation of this policy would be a disaster and that it represented serious fallacies about evaluation rigor. A portion of this letter stated:

> …we believe the proposed priority manifests fundamental misunderstandings about (1) the types of studies capable of determining causality, (2) the methods capable of achieving scientific rigor, and (3) the types of studies that support policy and program decisions. We would like to help avoid the political, ethical, and financial disaster that could well attend implementation of the proposed priority (Donaldson et al. 2009, p. 5).

These sentiments were not expressed by AEA's governing body alone. Other evaluators including those within the education community expressed dissatisfaction with the policy. One of them was Madhabi Chatterji from Columbia University's Teachers College. Chatterji (2007, p. 240) wrote:

> the federal "gold standard" for generating rigorous evidence on social programs and policies is not tenable with evaluative studies of complex field interventions such as those found in education. It introduces instead the concept of grades of evidence, illustrating

how the choice of research designs coupled with the rigor with which they can be executed under field conditions, affects evidence quality progressively. It argues that evidence from effectiveness research should be graded on different design dimensions, accounting for conceptualization and execution aspects of a study. Well implemented, phased designs using multiple research methods carry the highest potential to yield the best grade of evidence on effects of complex, field interventions.

There were evaluators within and outside of AEA who disagreed with the official AEA statement. In their unofficial statement, these evaluators emphasized the value of randomized control trials:

> Randomized controlled trials have been essential to understanding what works, what does not work, and what is harmful among interventions in many other areas of public policy including health and medicine, mental health, criminal justice, employment, and welfare. Furthermore, attempts to draw conclusions about intervention effects based on nonrandomized trials have often led to misleading results in these fields and there is no reason to expect this to be untrue in the social and education fields (Donaldson et al. 2009, pp. 9–10).

The unofficial statement also noted that the limitations of RCTs have been overblown by critics:

> Randomized trials in the social sector are more frequent and feasible than many critics acknowledge and their number is increasing. The Campbell Collaboration of Social, Psychological, Educational, and Criminological Trials Register includes nearly 13,000 such trials, and the development of this register is still in its youth. At the same time, we recognize that randomized trials are not feasible or ethical at times. In such circumstances, quasi-experimental or other designs may be appropriate alternatives, as the proposed priority allows. However, it has been possible to configure practical and ethical experimental designs in such complex and sensitive areas of study as pregnancy prevention programs, police handling of domestic violence, and prevention of substance abuse. It is similarly possible to design randomized trials or strong quasi-experiments to be ethical and feasible for many educational programs. In such cases, we believe the Secretary's proposed priority gives proper guidance for attaining high methodological standards and we believe the nation's children deserve to have educational programs of demonstrated effectiveness as determined by the most scientifically credible methods available (Donaldson et al. 2009, pp. 9–10).

Conclusion: The Need for an Integrated Approach

In spite of the fact that the wars may not be entirely over, many evaluators have attempted to move things forward by acknowledging the merit of both quantitative and qualitative approaches and advocating for "mixed-methods" (Hughes and Hutchings (2011)). Many recognize that both paradigms are needed for understanding both sides of humankind's nature: the physical world and the intellectual/ moral world (Lincoln 1991).

In complex programs, there are questions and situations that require either or a combinations of paradigms. Scriven in Smith (1994, p. 216) rates this question "right up there with thinking that the great problem for the U. S. public service is

whether it should hire men or women. The answer is yes! Hire one or the other or both, depending on the person's particular skills/competencies for the job at hand". Thus the key issue that evaluators need to master is: to "discern" which method is the most appropriate for the question and circumstance at hand (Greene et al. 2007). That is the gold standard in evaluation; proper choice of methods in terms of a) appropriate choice of research design(s) vis-à-vis the questions asked and b) how well the chosen design(s) can be and are executed under field conditions, given the constraints faced by researcher(s) (Chatterji 2007).

To embrace an integrated approach, evaluators also need to question implicit assumptions, sometimes the misconceptions, held about different paradigms. For example, the misconception is held that science is only about method and rigor. But science in evaluation is not limited to method and rigor. There are several other components of science in evaluation that evaluators have not been so concerned about-or thought to be forms of art and thus outside the definition of rigor. Lincoln (1991) outlines four such aspects:

- The science of locating interested stakeholders.
- The science of generating for them good, usable information to those stakeholders.
- The science of teaching various stakeholder groups how to use information to empower themselves.
- The science of communicating results.

Further, evaluators need to appreciate that the "art" in program evaluation are not just limited to what has been known as the "messy judgment ineffable". Other aspects of "art" need to be emphasized, such as those outlined by Lincoln (1991)

1. The art of not just our judgments as evaluators, but that of helping stakeholders to form their own informed and balanced judgments within program processes.
2. The art of appreciation—comprehending meaning within a context, understanding the social and cultural milieu from which a program draws its particular expression, seeing something in its wholeness and not in its pieces.
3. The art of listening for the deep meanings and not just search for responses to questions or whether objectives have been achieved.
4. The art of negotiating the worlds in which stakeholders live.
5. The art of dealing with people very different from ourselves, what programs mean, what evaluations mean to various stakeholders at different levels.

References

Association for the Advancement of Science (AAS), (1990). The Nature of Science. Retrieved 12/12/2011 from www.project2061.org
Anderson, P. (1999). Complexity theory and organization science. *Organization Science, 10*(3), 216–232.

Ayala, F. (1994). On the scientific method, its practice and pitfalls. *History and Philosophy of Life Sciences, 16*(1), 205–240.

Ball, S. J. (1995). Intellectuals or Technicians? The urgent role of theory in educational studies. *British Journal of Educational Studies, 43*(3), 255–271.

Chatterji, M. (2007). Grades of evidence: Variability in quality of findings in effectiveness studies of complex field interventions. *American Journal of Evaluation, 28*(3), 239–255.

Chen, H. T., & Garbe, P. (2011). Assessing program outcomes from the bottom-up approach: An innovative perspective to outcome evaluation. *New Directions for Evaluation, 2011*(30), 93–106.

Chelimsky, E. (2012). Valuing, evaluation methods, and the politicization of the evaluation process. In G. Julnes (Ed.) *Promoting valuation in the public interest: Informing policies for judging value in evaluation New Directions for Evaluation* (133(Spring), pp. 77–83).

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

Dattu, L. E. (1994). Paradigm wars: A basis for peaceful coexistence and beyond. *New Directions for Evaluation, 61*(Spring), 61–70.

Desrosieres, A (1998). *The politics of large numbers. A history of statistical reasoning*. Cambridge, MA: Harvard University Press.

Donaldson, I. S., Christie, C. A., & Mark, M. M. (2009). *What counts as credible evidence in applied research and evaluation practice?*. Los Angeles: Sage.

Dunn, W. N. (1998). Campbell's experimenting society: Prospect and retrospect. In W. N. Dunn (Ed.) *The experimenting society: Essays in honor of Donald T. Campbell* (pp. 20–21). New Brunswick, NJ: Transaction Publishers.

Dupre, J. (2001). *Human Nature and the Limits of Science*. *Simplification versus an extension of clarity!* Oxford: Clarendon Press.

Eade, D. (2003). *Development Methods and Approaches: Critical Reflections. A Development in Practice reader*. London: Oxfam GB.

Handa, S., & Maluccio, J. A. (2010). Matching the gold standard: Comparing experimental and nonexperimental evaluation techniques for a geographically targeted program. *Economic Development and Cultural Change, 58*(3), 415–447.

Hughes, K. & Hutchings, C. (2011). *Can we obtain the required rigor without randomization? Oxfam GB's non-experimental Global Performance Framework.*

International Initiative for Impact Evaluation Working paper 13. Retrieved October 10, 2011, from www.3ieimpact.org

House, E. R. (1984). Factional disputes in evaluation. *American Journal of Evaluation, 5*(19), 19–21.

Greene, J. C., Lipsey, M. W., & Schwandt, T. A. (2007). Method choice: Five discussant commentaries. *New Directions for Evaluation, 113*(Spring), 111–118.

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

Lay, M., & Papadopoulos, I. (2007). An exploration of fourth generation evaluation in practice. *Evaluation, 13*(4), 495–504.

Lincoln, Y. S. (1991). The arts and sciences of program evaluation. *Evaluation Practice, 12*(1), l–7.

Julnes, G. (2012a). Editor's note. *New Directions For Evaluation, 133*(Spring), 1–2.

Julnes, G. (2012b). Managing valuation. *New Directions for Evaluation, 133*(Spring), 3–15.

Murphy, N., Ellis, G. F. R., & O'Connor, T. (Eds.). (2009). *Downward causation and the neurobiology of free will*. Berlin: Springer.

Newman, J., Rawlings, L., & Gertler, P. (1994). Using randomized control design in evaluating social sector programs in developing countries. *The World Bank Research Observer, 9*(2), 181–201.

Nowotny, H. (2005). *Theory, culture & society*. London: Sage.

Roberts, A. (2002). A principled complementarity of method: In defence of methodological eclecticism and the qualitative-quantitative debate. *The Qualitative Report*, 7(3). Retrieved on July 3, 2011 from http://www.nova.edu/ssss/QR/QR7-3/roberts.html

Rowlands, J. (2003) Beyond the comfort zone: some issues, questions, and challenges in thinking about development approaches and methods. In E. Deborah (Ed.) *Development methods and approaches: Critical reflections. A development in practice reader* (pp. 1–20). London: Oxfam GB.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.

Shadish, W. R., Cook, T. D. & Leviton, L. C. (1991). Donald T. Campbell: Methodologist of the experimenting society. In W. R. Shadish, T. D. Cook, & L. C. Leviton (Eds.), *Foundations of program evaluation* (pp. 73–119). London: Sage.

Smith, M. F. (1994). On past, present and future assessments of the field of evaluation. *Evaluation Practice*, *15*(3).

Smith, N. L. (2010). Characterizing the evaluand in evaluating theory. *American Journal of Evaluation, 31*(3), 383–389.

Social Sector Programs in Developing Countries (1994). *The World Bank Research Observer*, *9*(2), 181–201.

Toulmin, S. (2001). *Return to reason*. Cambridge, MA: Harvard University Press.

Wrigley, T. (2004a). School effectiveness': The problem of reductionism. *British Educational Research Journal, 30*(2), 227–244.

Warren, A. (2011) *The myth of the plan*. Retrieved December 08, 2011, from http://stayingfortea.org/2011/06/27/the-myth-of-the-plan/

Wrigley, T. (2004b). School effectiveness': The problem of reductionism. *British Educational Research Journal, 30*(2), 227–244.

Zhu, S. (1999). A method to obtain a randomized control group where it seems impossible. *Evaluation Review, 23*(4), 363–377.

Part II
Evaluation Theory and Assumptions

Chapter 5
Theory in Evaluation

Abstract Distinctions are drawn among social science theory; evaluation theory; evaluators' theory; and program theory. Similarly, the distinguishing characteristics that mark differences between results chains and theories of change are discussed. The chapter underlines the concept that logic models are incomplete if they do not spell out the assumptions (implicit/or behind-the-scenes concepts, conditions, and qualifications) within the linkages or arrows in a results chain, which explain why result A is expected to lead to result B and so on.

Introduction

How we design and conduct evaluations is much influenced by why we do them. Evaluations are undertaken for many different reasons. At the broadest level, these are reflected in the very ways evaluation is defined in the context in which the evaluator works. Some definitions emphasize the role of evaluation as "judging the worth or merit" of an object. e.g. Organization for Economic Cooperation in Development (OECD) Development Assistance Committee (DAC) 2002:13 says evaluation is "the process of determining the worth or significance of an activity, policy or program." Sometimes evaluation definitions emphasize the "usefulness" of the feedback about the evaluand (e.g. Trochim 2000-defines evaluation as "An assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention").

According to Weiss (1972) the purpose of evaluation research is to measure the effects of a program against its goals as a means of improving subsequent program decision making and improving future programming. Weiss emphasizes four features that characterize evaluation purpose: (a) to measure the effect, which refers to the research methodology that is to be used; (b) the effects—the outcomes

of the program, rather than its efficiency, honesty, morale, or adherence to rules or standards; (c) the comparison of effects with goals, which stresses the use of explicit criteria for judging how well the program is doing; (d) the contribution to subsequent decision making and/or the improvement of future programming. These embody the "social" purpose of evaluation.

Weiss (1993, pp. 93–94) argues in a later work that evaluation research is a rational enterprise that examines the effects of policies and programs on their targets—whether individuals, groups, institutions, or communities—in terms of the goals these initiatives are meant to achieve. By objective and systematic methods, evaluation research assesses the extent to which goals are realized and looks at the factors that are associated with successful or unsuccessful outcomes. It is expected that careful and unbiased data on the consequences of programs will improve decision-making. According to Fitzpatrick et al. (2010) the primary purpose of the evaluation is "to develop a good understanding of, and to make some judgments about, the quality and effectiveness of the program". This definition emphasizes understanding and program improvement. Her purpose is formative rather than summative, and her primary audience for this formative information is the program staff.

Whether the emphasis of the evaluation is on judging worth or generating feedback, stakeholders will need to know if the program is delivering its intended results. Often, programs are funded and implemented with the belief that the program is likely—but not guaranteed to bring about the intended results, in light of a myriad of other influences (Mayne 2011). Program managers always need to figure out the extent of change within the program's target groups, and whether it is reasonable to draw a dotted line between the program and the change. Hendricks (1966) in Mayne (2011) uses the term 'plausible association' to refer to this link.

It is increasingly recognized that "theory" is an important part of making evaluations effective e.g. when there is need for 'plausible association'. However, the term "theory" is commonly used in program evaluation to mean a range of different things. It may reference general theory or propositions, social science theory, evaluation theory, or program theory.

General Theory

Theories are sets of propositions that jointly provide explanations and integration (Scriven 1998). In science, a theory is a more-or-less established explanation for some phenomenon (Smith 2010). Theories are aimed at conceptual simplification and/or explanation of the phenomena; they involve a greater degree of speculation than does direct observation. Prediction is just a bonus, not a necessary condition for something to be a theory, even a good theory (Scriven 1998, p. 57). Some theories try to provide explanations and or an 'account' rather than a 'theory' of

the field. Such exercises are sometimes called "philosophies" of the field rather than theories in the strict sense. Scriven emphasizes that a theory is not merely an outline of components. There should also be an explanation of how the components fit together, which Scriven (1998, p. 57) refers to as "the unifying logic of operation".

Social Science Theory

Rather than focusing on methods or practices of evaluators, social science theories attempt to provide generalizable and verifiable knowledge about the principles that shape social behavior (Donaldson and Lipsey 2006).

According to Fay in Ball (1995) the role of theory is not limited to describing "social order". It also influences it. "Social theory does not simply offer a picture of the way that a social order works. Instead, a social theory itself is a catalytic agent of change within a complex of social life which it analyses". (p. 263). This may in many instances be true, since policy makers are very likely to draw on their banks of "theory" and "beliefs" when making decisions that influence social order. It is also worth noting in the same vein that social order influences theory, since ongoing policy making and implementation generate new formations of theory.

Fay also argues that theory is a vehicle for reflective thinking.

> *Theory is a vehicle for thinking otherwise; it is a platform for outrageous hypotheses and for unleashing criticism… Theory is destructive, disruptive and violent. It offers a language for challenge and modes of thought other than those articulated for us by dominant others. It provides a language of rigor and irony, rather than contingency (Fay in* Ball 1995, p. 267).

Reflective thinking, as proposed by Fay, is similar to policy implementation in the sense that both processes test the validity of theories and generate new theory formations. However, in Fay's "reflective thinking", theory is the very forum in which one set of ideas is challenged by another. This may provide an avenue of "refining" theories before they can be tested, e.g., in the form of policy implementation. It is worth noting, however, that "reflective thinking" may not be so much a function of theory per se as it is a function of theorizing, a discourse in which theorists or practitioners or policy makers (Fay refers to these as 'policy engineers') generate, critique, and test theories. In this case, the function of critical thinking may be a quality of the theorizing discourse, which depends on the specific parties involved and how they interact. Fay warns of the risk that certain parties may subdue others, thereby biasing the discourse; a folly he refers to as an "assertive recycling of dogmas and tired utopias" (Fay in Ball 1995, p. 267).

Evaluation Theory

According to Alkin 2004, (cited in Donaldson and Lipsey 2006, p. 17) evaluation theories are largely prescriptive and offer a set of rules, prescriptions, prohibitions, and guiding frameworks that specify what a good or bad evaluation is and how evaluation should be done.

Theories are key to evaluation. Coryn et al. (2011) identify two roles of evaluation theories:

(a) Evaluation theories describe and prescribe what evaluators do or should do when conducting evaluations.
(b) They specify such things as purpose, users, and use. Additionally, they identify who participates in the evaluation process and to what extent, general activities or strategies, method, choices, and roles and responsibilities of the evaluator, among others.

Shadish (1998) in Donaldson and Lipsey (2006, p. 18) identifies other six reasons why evaluation theory is important. It is useful in:

(a) Providing a language that evaluators can use to talk with each other about evaluation.
(b) Encompassing many things in our field about which evaluators seem to care most deeply.
(c) Defining the themes of the majority of evaluation professional conferences.
(d) Providing evaluators with an identity that is different from the identity of other professionals.
(e) Providing the face that evaluators present to the outside world.
(f) Comprising the knowledge base that defines the evaluation profession.

Evaluation theory is different from the evaluator's theory, which Greene et al. (2007) defines thus: "this complex constellation of assumptions, expertise, values, and judgment can be called an evaluator's theory—his or her understanding of the social practice of evaluation". These theories are informed by a blend of academic study, disciplinary socialization,and professional experience. Evaluator theory is important because, it influences decisions about evaluation methodology, since it is on basis of the evaluator's theory that she or he responds to issues presented and makes judgments for the most viable options within the context.

Program Theory

Lipsey (1987) in Chen (1990, p. 7) defines program theory as "a set of propositions regarding what goes on in the black box during the transformation of input into output;" that is, how, "via treatment inputs, a bad situation is transformed into a better one".

Program theories specify relationships among program actions, outcomes, and other factors (Coryn et al. 2011). Activities are the actions undertaken to bring about a desired end; outputs are the immediate result of an action; outcomes are the anticipated changes that occur directly or indirectly as a result of inputs, activities, and outputs (Coryn et al. 2011). Initial outcomes are usually expressed as changes in knowledge, skills, abilities, and other characteristics (e.g., increased knowledge of safe sexual practices). Intermediate outcomes are often classified as behavioral changes (e.g., increased use of condoms) that are believed to eventually produce changes in long-term outcomes, such as the alleviation, reduction, or prevention of specific social problems or meeting specified needs of a program's target population (Coryn et al. 2011). Leeuw (2003) points out a distinction between two commonly used terms—program theory and program logic. While program logic often specifies the inputs and components of a program, as well as short-term and long-term outcomes, along with the assumed linkages among these; program logic rarely outlines the underlying mechanisms that are presumed to be responsible for those linkages.

Chen (2005) defines program theory as "a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated." This definition entails two parts:

(a) Prescriptive/normative theory: what the structure of the program should be, including the intervention, implementation processes, and outcomes
(b) Descriptive/causative theory: the underlying construct of causal mechanisms that link the relationships among program treatments, implementation processes, and outcomes.

Normative theory provides guidance on what goals or outcomes should be pursued or examined and how the treatment should be designed and implemented. It can come from unexamined premises, assumptions, procedures, or prior knowledge and theory, or all of these. It is usually taken for granted by program designers or other stakeholders, and thus not explicit or systematic. Yet it provides the rationale for the program structure and activities. Normative theory provides program planning, formulation and implementation grounding. An examination of normative theory is important, as it helps program designers and teams better understand the conceptualization and assumptions, and also think through crucial issues concerning program implementation. When the intervention is designed and implemented appropriately, the credibility of the program is enhanced. Thus, although it is important to provide timely information for improving program activities, the evaluation of a normative theory alone is not enough to understand whether the program achieved its intended goals or to identify the unintended consequences of the program. Hence, the need for causative theory.

Causative theory specifies how the program works by identifying the conditions under which certain processes will arise and what their likely consequences will be. It is usually empirically based. It represents the empirical knowledge about the relationship between the intervention and the outcome, or at least interventions and

outcomes seen as closely similar. Causative theory provides information on the impacts generated by a program and how the impacts are produced. Causative theory is useful in providing information on the contextual and intervening factors that affect a particular program's development.

As Stame (2004) argues, the traditional black box approach to program evaluation simply provides information on whether a program failed or succeeded. If the program failed, the approach is limited in explaining why the program failed or what could have been the major weakness of the program. When the causative theory is mapped out elaborately, the evaluation can easily point out the weaknesses in the causal mechanism that underlies the program and identify the contextual and intervening factors that hinder or facilitate program processes and outcomes. It can also suggest possible strategies for dealing with these issues.

According to Chen (2005), the major domains of normative theory include:

Treatment domain: The treatment is the essential element (or collection of elements) within the program that cause(s) the desired change. Treatment theory specifies what the nature of program's treatment should be.

Implementation environment domain: The implementation environment—in which the intervention is implemented–affects the program processes and consequences. It concerns itself with whether the intervention is reaching the target group to which evaluation findings will be generalized or even applied. Implementation environment theory specifies the nature of the contextual environment within which the program should be implemented.

Outcome domain: This domain defines which outcomes should be examined. Outcome theory specifies what the nature of the program outcome should be.

According to Chen (2005), the domains related to causative theory include:

Impact domain: assesses the relationship between the treatment and outcome domains, and specifies the causal relationship between the intervention and the results.

Generalization domain: the extent to which processes and outcomes of a research system can be applied to a general system. Generalization theory specifies the generalizability of evaluation findings to the topics or circumstances of interest to stakeholders.

Intervening mechanism domain: the intervening processes that link treatment domain with outcome domains. Intervening mechanism theory specifies how the underlying intervening processes operate. According to Hansen and Vedung (2010), intervention theories are defined as presuppositions (notions, conceptions, and assumptions) of how an intervention (a program, a policy, a treatment, and/or an organizational change) may affect a given situation to change it or preserve it in ways that are preferable or not preferable to the situation absent the intervention or with different intervention. Hansen and Vedung (2010) suggest that an intervention theory is composed of three elements:

- *Situation theory*: Notions concerning relevant features of the context in which an intervention is supposed to take place, including (a) size of the problem at present; (b) development of the problem up until the present time; (c) size of the

problem after "x" years if nothing is done; (d) causes of the problem, and finally (e) impact of the problem at present and after x years if nothing is done.

- *Causal theory*: Notions concerning how a given intervention/program directly or indirectly, through its processes of implementation and delivery of outputs will (a) have an impact on the causes of the underlying problem at issue so that the problem will disappear, be reduced, or be prevented from becoming aggravated and (b) have possible effects in other areas during the entire process such as people outside the target group.
- *Normative theory*: Notions concerning why the various aspects of the underlying situation or problem that are supposed to be affected by the intervention are preferable or not preferable to the situation without the intervention or with different intervention.

Theory-Based Evaluation (TBE)

Theory-based evaluation has its origins in the work of Tyler in the 1930s (with his notion of formulating and testing program theory for evaluative purposes). The approach has been popularized by Chen, Weiss, Suchman, Rossi, Rogers, Petrosino, Huebner, and others (Coryn et al. 2011). The approach is based on careful articulation of the program theory or models and the use of these theories/models as a guiding framework for evaluation. It sets out the theoretical assumptions underlying an intervention in terms of a phased sequence of causes and effects—a program theory (Carvalho and White 2004; Birckmayer and Weiss 2000). Theory-based evaluation involves identifying key service components and expected program outcomes, and working with programs to make explicit the underlying assumptions about how these service components will lead to the desired outcomes. These services, outcomes, and the hypothesized links between them are the basis for developing a program model or theory. This program theory becomes the framework to guide the development, implementation, and interpretation of the evaluation (Green and McAllister 1998, p. 4). This may involve asking what activities are being conducted, what effect each particular activity will have, what the program does next, what the expected response is, what happens next, and so on, reaching all the way to the expected outcomes (Birckmayer and Weiss 2000).

Within the framework of program theory, data are then collected to examine how well each step of the sequence is realized. Data analysis attempts to determine whether the required conditions are in place for the desired outcome. If they are, and if the theory is correct, chances are that the program was successful. Coryn et al. (2011) outline the principles and subprinciples of theory-based evaluation (see Box 5.1).

Box 5.1: Core Principles and Subprinciples of Theory-Driven Evaluation

1. Theory-driven evaluations/evaluators should formulate a plausible program theory

   a. Formulate program theory from existing theory and research (e.g., social science theory)
   b. Formulate program theory from implicit theory (e.g., stakeholder theory)
   c. Formulate program theory from observation of the program in operation/exploratory research (e.g., emergent theory)
   d. Formulate program theory from a combination of any of the above (i.e., mixed/integrated theory)

2. Theory-driven evaluations/evaluators should formulate and prioritize evaluation questions around a program theory

   a. Formulate evaluation questions around program theory
   b. Prioritize evaluation questions

3. Program theory should be used to guide planning, design, and execution of the evaluation under consideration of relevant contingencies

   a. Design, plan, and conduct evaluation around a plausible program theory
   b. Design, plan, and conduct evaluation considering relevant contingencies (e.g., time, budget, and use)
   c. Determine whether evaluation is to be tailored (i.e., only part of the program theory) or comprehensive

4. Theory-driven evaluations/evaluators should measure constructs postulated in program theory

   a. Measure process constructs postulated in program theory
   b. Measure outcome constructs postulated in program theory
   c. Measure contextual constructs postulated in program theory

5. Theory-driven evaluations/evaluators should identify breakdowns, side effects, determine program effectiveness (or efficacy), and explain cause-and-effect associations between theoretical constructs

   a. Identify breakdowns, if they exist (e.g., poor implementation, unsuitable context, and theory failure)
   b. Identify anticipated (and unanticipated), unintended outcomes (both positive and negative) not postulated by program theory

    c. Describe cause-and-effect associations between theoretical constructs (i.e., causal description)

    d. Explain cause-and-effect associations between theoretical constructs (i.e., causal explanation)

      i. Explain differences in direction and/or strength of relationship between program and outcomes attributable to moderating factors/variables

      ii. Explain the extent to which one construct (e.g., intermediate outcome) accounts for/mediates the relationship between other constructs

Source: Coryn et al. (2011)

An example of how a theory-based evaluation may be conceptualized is illustrated by the following example of evaluation questions that may be generated for a counseling program (Smith 2010, pp. 383–389):

> Does the program succeed in matching counselor to students by race/ethnicity, and if so, do students see the counselors as models to emulate? Do the students see the counselors more as "big brothers/big sisters" than, say, as teachers or authority figures, and if so, does the nature of the relationship bring about a sense of trust? Do the counselors implement the suggested set of learning activities in the order planned, and if so, do students experience success in moving through the sequence? Do they gain confidence in their academic abilities? Does the program provide the planned kinds of field trips, and do students gain from the experience a better knowledge of possible educational and vocational futures? Do students have greater motivation to work hard at their school work? The outcome question will be the following: Do students, in fact, do better in school? Any evaluations would ask that question and collect relevant data. What TBE does, in addition, is to track the steps along the route. It seeks to find out whether the theories on which the program is based are realized in action

In support of theory-based evaluation, Chen (2005) argues that tracing developments in mini-steps, from one phase to the next, helps to ensure that the evaluation is focusing on real effects of the real program and that the often-unspoken assumptions hidden within the program are surfaced and tested. Stame (2004, p. 59) argues that theory-based evaluation provide a way of dealing with two problems that have plagued evaluation—the black box and evaluation deficit. According to Stame (2004, p. 59):

> …the black box is the space between the actual input and the expected output of a program. Moved by the need to tackle serious social problems, program designers often gloss over what is expected to happen, the how and why, when an input is put in place; and evaluations do the same concentrating on measuring outputs, while attributing the observed difference to the input. All this is hardly informative for a policy design wishing to build upon previous experience. However, for a long time evaluations have coexisted with black box programs, and have tried to cope with similar shortcomings by developing sophisticated methods for measuring the distance between objectives and results.

According to Stame (2004), the evaluation deficit refers to the unsatisfactory situation in which most evaluations, conducted at local and other sub-national levels, provide the kind of information (on output) that does not immediately

inform an analysis of effects and impacts at higher levels, i.e. whether global objectives have been met. Or conversely, impact assessment is not corroborated by an understanding of the working of programs. Consequently, evaluations are under-utilized or not utilized at all. Although this has not been diagnosed as a black box problem, the vertical predicament shares many points in common: generalizations are not easily drawn, and it is not clear how lessons learned at the local level can be useful at higher levels. There is an even more striking similarity: the issue keeps being tackled as though it were either a lack of government capacity or of bad quality evaluation methods.

Stame argues that theory-based approaches may have helped open the black box of "complex" programs at a community level, but programs enacted through multi-level governance seem to suffer from a double kind of evaluation deficit: on the one side, under-utilization of lower level evaluations and, on the other, impact assessments unable to offer an explanation of why and how outcomes occurred.

Theory-based evaluation approaches offer some clues to address this set of problems. Considering the beneficiaries of programs as actors situated in a stratified reality in which they will act and react, theory-based approaches allow for different ways of conceiving how particular means and policy instruments will work to produce good outcomes. The latter should not therefore be seen as the result of the homogeneous implementation of some practice, nor should they be added up, to establish who fared well, who fared average, and who fared the worst. Instead, the higher policy level should have the ability to recognize the potentialities of the lower levels, and of admitting to a multiplicity of means to be adapted to any given situation (Stame 2004, p. 59).

Theory-based evaluation offers a number of variations in its approaches to programs. Stame (2004) notes a distinction between theory-driven evaluation (by Chen and Rossi) and theory-based evaluation (by Weiss). The main tenet of theory-driven evaluation is that black-box programs are such because they have no theory, goals are unclear, and measures are false, with the result that evaluations are "at best social accounting studies that enumerate clients, describe programs and sometimes count outcomes." That is, "The black box is an empty box" (Stame 2004). The emphasis of theory-driven evaluation then is more than just to provide a program's missing theory. On the other hand, theory-based evaluation emphasizes the role of politics. Good program theory is not theory that is unaffected by politics, but a good theory of how politics works in practice. In this connection, Scriven (1998) argues for another variation that has come to be known as "Theory-based Stakeholder Evaluation (TSE)." According to Hansen and Vedung (2010), TSE differs from other theory-based evaluation approaches in this way: while all of the others fuse the program perceptions of the various stakeholder groups into one unitary program theory; the TSE model keeps the program theories of the diverse stakeholder groups apart from each other and from the program theory embedded in the institutionalized intervention itself. This is expected to represent an important clarification and extension of the standard theory-based evaluation. Although these variations exist, theory-based evaluation approaches continue to be known for their emphasis on elaborate program theory or theory of change (ToC).

Some evaluators have expressed sentiments that theory-based evaluation approaches have been over-emphasized. For example, Scriven (1998) argues, in reference to Chen's definition of program theory (an account of how the program produces outputs and outcomes emerge from outputs, as well as side effects), that one does not need a theory to know or predict that certain outcomes will emerge from a process, just like one may not need a theory of how aspirin works but still predict its effects.

Also, Scriven argues that there are circumstances where theory-based evaluation may not be the most appropriate:

> … people sometimes think that all evaluation should be theory-driven, that this is the natural way to do evaluation. That shows a lack of understanding of the necessity to do black-box evaluation when no theory is available or needed, or when getting the relevant expertise would take resources of time or money we do not have. It is a very practical point to avoid doing what is expensive and not needed; and a little understanding of the general nature of evaluation will avoid this error of practice (Scriven 1998, p. 57).

Scriven (1998) also cautions that identifying program components alone is not theory-driven evaluation, as this should happen in any analytic evaluation. Demonstrating how the pieces are linked together is an essential part of theory-based evaluation. Astbury and Leeuw (2010) have used the term "mechanisms" as opposed to "program activity" to conceptualize how programs work. They refer to "mechanisms" as "unexplained "causal arrows" that seem to flourish so well in the present climate of enthusiasm with visual logic models" (Astbury and Leeuw 2010, p. 363). They characterize "mechanisms" with three major features:

(a) Mechanisms are not independent causal or intervening variables that attempt to account for why a statistical correlation exists between an independent and dependent, rather, they are explanations of why variables are related.
(b) Mechanisms are usually hidden/underlying and hence often unobservable or ''hidden.'' Thus, in order to explain program outcomes, evaluators should go "beneath" the patterns of data to explain factors that may account for those patterns. Mechanisms are sensitive to variations in context: some mechanisms may be inactive in certain contexts while dominant in others.
(c) Mechanisms generate outcomes; thus causation cannot be directly observed—only inferred—by examining patterns of regular contingent relations between events.

The other term that has been used to conceptualize "how programs work" is theories of change, which has been discussed in Chap. 3 (program frameworks). Theories of change have often used synonymously with program theory, but they often highlight an emphasis on underlying assumptions about causal linkages within the program and other factors that should influence the program's success. Mason and Barnes (2007) argue that a crucial role of theories of change is to help in determining plausible associations. A theory (of change) is necessary in order to unpack the 'black box' that lies between project interventions and outcomes and to be precise about what it is within projects that leads to (intended or unintended)

outcomes. According to Mayne (2011), a results chain (or logic model) per se is not a ToC. The ToC spells out the assumptions behind the arrows in a results chain—the behind-the-scene assumptions or why result A is expected to lead to result B. (Mayne 2011; Mercieret al. 2000). Often when funders require the development of a program logic model, the emphasis is on laying out the activities and expected outcomes. Although articulating the underlying rationale for a program is critical to its success, it is frequently a second-generation or post hoc activity—one that is never quite completed. In addition, the discussion of program assumptions seems to be the place where there is the biggest disconnect between planners/evaluators and program managers/implementers. Often managers are willing and able to lay out activities and expected outputs, but the time-consuming process of articulating and assessing the strength of assumptions through a literature review or discussion with experts can feel like a distraction. According to Aspen Institute's Roundtable on Comprehensive Community Initiatives (RTCCI) (Anderson 2004), a ToC is essentially an explanation of how a group of stakeholders expects to reach a commonly understood long-term goal. The TOC process hinges on defining all of the necessary and sufficient preconditions required to bring about a given long-term outcome. RTCCI's ToC model comprises (a) a pathway of change that illustrates the relationship between a variety of outcomes that are each thought of as preconditions of the long-term goal; (b) indicators that are specific enough to measure success; (c) interventions that are used to bring about each of the preconditions on the pathway and at each step of the pathway; and (d) assumptions that explain why the whole theory makes sense!

The ToC approach has been popularized as an approach for designing and evaluating complex programs. One of the evaluation approaches that have popularized the importance of ToC is contribution analysis. This type of analysis was developed by John Mayne for situations where designing an 'experiment' to test cause and effect is impractical (Mayne 2001, 2011; Hendricks 1996).Contribution analysis involves explicitly exploring the most likely alternative explanations, presenting evidence to discuss them, and where appropriate, discounting them. This helps to reduce the uncertainty about the contribution made and strengthens the argument in favor of the program's impact (Hendricks 1996).

Developing a postulated ToC is a key step in contribution analysis (see Box 5.2).

Box 5.2: Steps in Contribution Analysis

Step 1: Set out the cause-effect issue to be addressed

- Acknowledge the attribution problem.
- Scope the problem: determine the specific cause-effect question being addressed; determine the level of confidence needed in answering the question

- Explore the contribution expectations
- Determine the other key influencing factors
- Assess the plausibility of the expected contribution given the program size
  Step 2: Develop the postulated theory of change and risks to it

- Set out the postulated theory of change of the program, including: identify the roles of the other influencing factors in the theory of change, identifying risks to the assumptions in the theory of change, and determine how contested is the postulated theory of change
  Step 3: Gather the existing evidence on the theory of change

- Assess the strengths and weaknesses of the links in the embedded results chain
- Gather the evidence that exists from previous measurement, past evaluations, and relevant research (1) for the observed results (2) for each of the links in the results chain, and (3) for the other influencing factors
  Step 4: Assemble and assess the contribution story, and challenges to it

- Set out the contribution story
- Assess the strengths and weaknesses in the postulated theory of change in light of the available evidence, and the relevance of the other influencing factors
  Step 5: Seek out additional evidence

- Determine what kind of additional evidence is needed to enhance the credibility of the contribution story.
- Refine the theory of change
- Gather new evidence
  Step 6: Revise and strengthen the contribution story

- Build the more credible contribution story
- Reassess its strengths and weaknesses
- Revisit step 5
  Step 7: In complex settings, assemble and assess the complex contribution story

- Develop the contribution story for each additional sub-theory of change (using steps 2–6)
- Develop the contribution story for any general theory of change (using steps 2–6)

Adapted from Mayne (2011)

According to Merriam Webster's dictionary, to postulate is to assume or claim as true.

Synonyms include hypothesize, assume, premise, presume, presuppose, say, or suppose. A postulated theory is what program stakeholders expect to happen with the implementation of the program. It should set out the steps in the results chain and in addition to laying out these steps; identify the assumptions behind the various links in the results chain and the risks to those assumptions (Mayne 2011). Underlying assumptions are made explicit in theory-based evaluations where evaluation tracks systematically the anticipated results chain. Assumptions of their nature and usefulness, are the focus of the next two chapters.

References

Anderson, A. (2004). *Theory of change as a tool for strategic planning: a report on early experiences*. Washington: Aspen Institute.

Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation, 31*(3), 363–381.

Ball, S. J. (1995). Intellectuals or technicians? The urgent role of theory in educational studies. *British Journal of Educational Studies, 43*(3), 255–271.

Birckmayer, J. D., & Weiss, C. H. (2000). Theory-based evaluation in practice what do we learn? *Evaluation Review, 24*(4), 407–431.

Carvalho, S., & White, H. (2004). Theory-based evaluation: the case of social funds. *American Journal of Evaluation, 25*(2), 141–160.

Chen, H. T. (2005). *Practical program evaluation: assessing and improving planning, implementation, and effectiveness*. Newbury Park, California, London and New Delhi: Sage publications.

Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schroter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 2*(2), 199–226.

Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in evaluation practice. In I. Shaw, J. Greene, & M. Mark (Eds.), *Handbook of evaluation*. Thousand Oaks: Sage.

Fitzpatrick, S., Pleace, N., & Wallace, A. (2010). *Evidence analysis for the triennial review: homelessness*. London: Equality and Human Rights Commission.

Greene, J. C., Lipsey, M. W., & Schwandt, T. A. (2007). Method choice: five discussant commentaries. *New Directions for Evaluation, 113*, 111–118. Spring.

Green, B. L., & McAllister, C. (1998). Theory-Based, Participatory Evaluation: A Powerful Tool for Evaluating Family Support Programs. *The Bulletin of the National Center for Zero to Three, Feb/March, 1998*, 30–36.

Hansen, M. B., & Vedung, E. (2010). *Theory-Based Stakeholder Evaluation American Journal of Evaluation, 31*(3), 295–313.

Hendricks, M. (1996, November). *Performance monitoring: how to measure effectively the results of our efforts*. Presented at the American Evaluation Association Annual Conference, Atlanta.

Leeuw, F. L. (2003). Reconstructing program theories: methods available and problems to be solved. *American Journal of Evaluation, 24*(1), 5–20.

Mason, P., & Barnes, M. (2007). Constructing theories of change: methods and sources. *Evaluation, 13*(2), 151–170.

Mayne, J. (2011). Contribution analysis: addressing cause effect. In K. Forss, M. Marra & R. Schwartz (Eds.), *Evaluating the Complex: Attribution, Contribution, and Beyond* (pp. 53-96). New Brunswick, New Jersey: Transactional Publishers.

Mercier, C., Piat, M., Peladeau, N., & Dagenais, C. (2000). An application of theory-driven evaluation to a drop-in youth center. *Evaluation Review, 1*, 73–91.

Scriven, M. (1998). Minimalist theory: the least theory that practice requires. *American Journal of Evaluation, 19*(1), 57–70.

Smith, N. L. (2010). Characterizing the evaluand in evaluating theory. *American Journal of Evaluation, 31*(3), 383–389.

Stame, N. (2004). Theory-based evaluation and types of complexity. *Evaluation, 10*(1), 58–76.

Trochim, W. (2000). *The research methods knowledge base* (2nd ed.). Cincinnati: Atomic Dog Publishing.

Weiss, C. H. (1972). *Evaluation research—methods for assessing program effectiveness*. Englewood Cliffs: Prentice Hall.

Weiss, C. H. (1993). Where politics and evaluation research meet. *American Journal of Evaluation, 14*(1), 93–106.

Chapter 6
What are Assumptions?

Abstract Although the pertinence of implicit assumptions is underlined in theory-based evaluations, the nature of these assumptions has been rarely discussed. To understand the nature of underlying assumptions, it is necessary to review the notion of assumptions in general and its remarkable role in the generation of knowledge. This chapter discusses the nature of assumptions and groups in 10 categories of assumptions according to Brookfield's (Becoming a critically reflective teacher. San Francisco, CA: Jossey-Bass, 1995) typology: Paradigmatic, Prescriptive and Causal assumptions and according to the degree of articulation.

> *Even the most well-intentioned person unwittingly allows unconscious thoughts and feelings to influence apparently objective decisions*
>
> Banaji, M. N. in Fine and Handelsman 2006, p. 1

> *Every human society rests on assumptions that, most of the time, are not only unchallenged but not even reflected upon. In other words, in every society there are patterns of thought that most people accept without question as being of the very nature of things*
>
> Trachman and Bluestone 2005, p. 131

> *To deny a proposition is not the same as to confirm its denial. …Given a proposition P, there is an associated proposition not-P. Either of these…may be merely supported or assumed. But when we deny P, we are not concerned with mere assumption, and there is nothing to be done with P that is logically equivalent to assuming not-P….the state of mind in which we reject a proposition is not the same as that in which we accept its negation*
>
> Russell 1904, p. 336

Introduction

Theory-based approaches emphasize the need to examine "underlying assumptions" in program theory. However, they do not do enough in examining the nature of these assumptions and methods of explicating them. The journey to understanding the nature of underlying assumptions has to start with an exploration of assumptions' roots in scientific inquiry.

Assumptions, underlying or explicit, are not only pertinent to program theory; they also have been a vital part of the evolution of knowledge. For example, Harste et al. (1984) have argued that "fact seeking", the goal of science, proceeds on countless assumptions that always need to be verified. These assumptions may be presented as beliefs, belief systems, or theoretical paradigms or they may not be articulated at all. If the assumptions are articulated, the pursuers of knowledge, then, gather evidence in light of these beliefs (Sosa 1974; Kanazawa 1988; Read 2003). When evidence shows that a belief is faulty, a new theory or belief statement emerges, hence a paradigm shift.

Paradigm shifts emerge when people (scientists or not) question their assumptions (tacit or explicit). Otto Van Neurath (in Harste et al. 1984, p. 85) has used a metaphor of shipbuilding to describe this process: "Scientists are like sailors who have to rebuild their ships at high sea, without being able to seek port. Each plank in the hull may be jettisoned in the process, but it is not feasible to jettison all of the planks at the same time". Thus, the "ships" of knowledge are built with pieces of wood (assumptions). The "worth" of each piece has to be examined from time-to-time. The pieces that are outdated by new information or disproven need to be discarded (jettisoned). Based on this metaphor, it can be said that "knowing" is about making and testing assumptions. Science is an activity of reflective thinking. Brookfield (1995) calls this hunting assumptions and Fogarty (1994) refer to this as meta cognition, reflecting on how we think, how we know, and how we learn.

But assumptions are not only crucial in the science realm; they are woven into our daily lives. We base all sorts of actions on them. A driver on the highway assumes that other drivers will obey traffic signals. When the driver proceeds through an intersection based on a green light in his favor, he assumes that the cross traffic will stop at its red light (Kies 1995). This can be a dangerous way to live our daily lives or to practice our professions.

Bennett's (1933, p. 157) story illustrates the potential danger when practice is based on assumptions. One evening the instructor began to discuss the solution of a problem. He first announced that "before I attempt the solution of this problem, I shall make an assumption." He then proceeded to illustrate and discuss his solution to the problem. He concluded with an invitation for questions. And one good student asked, "Having based your entire solution process upon an assumption; is it not true that your conclusion also must be an assumption?" This question, Bennett reported, was never answered to the satisfaction of the class. Although these students might have been discontented with the fact that the instructor's

solution was based on an assumption, the merit of that assumption itself may have been a greater source of contention. This is a case where parties consciously make assumptions and even weigh the merits of the assumption.

This conscious articulation of assumption(s) is not a frequent occurrence, however. Most often, assumptions are neither consciously made nor explicitly articulated. Thus, stakeholders may be completely unaware of the assumptions being made, whether these assumptions are in conflict, and whether they are even relevant to the known facts and conditions. We often hold tacit or unconscious assumptions that influence our judgments and expectations (Sherman and Harman 2011).

Much more specifically, it has been argued that development programs have often been based on such tacit assumptions. Kotze (1987, pp. 31–36) outlines some of the contradictory assumptions on which community development has long been based (See Box 6.1).

Box 6.1: Contradictions and Assumptions in Community Development (Kotze 1987, pp. 31–36)

Definitions of Assumptions

According to the Merriam-Webster dictionary (2011), an assumption is something that is taken for granted-or advanced as a fact. Synonyms include posit, postulate, premise, presumption, presupposition, to pretend to have or be, and something accepted as true. To assume is to suppose, to take it as given, to take for granted, take as read, conjecture, surmise, conclude, deduce, infer, reckon, reason, think, fancy, understand, gather, and figure. According to Kies (1995), assumptions are beliefs or ideas that we hold to be true—often with little or no evidence required. Brookfield (1995) defines assumptions as the taken for-granted beliefs about the world, and our place within it, beliefs that seem so obvious as not to need explicit statement. Implicit or tacit assumptions are often referred to as presumptions (Ullman-Margalit 1983).

According to Kanazawa (1988), assumptions are universal axiomatic statements about some part of the empirical world. An assumption set is a set of all assumptions for a given theory. Hypotheses are deduced from assumptions through logic alone without aid of any empirical knowledge. The only requirement for an assumption set is internal logical consistency; assumptions of a given theory may not logically contradict each other (Kanazawa 1988). The assumption set is the logical starting point of a causal explanation in scientific theory and produces a set of empirically testable hypotheses. A theory is tested by subjecting its hypotheses to empirical evidence (Kanazawa 1988).

Types of Assumptions

Brookfield (1995) classifies assumptions into three categories: paradigmatic, prescriptive, and causal.

Paradigmatic Assumptions

Paradigmatic assumptions are what we use to structure the world around us. A paradigmatic assumption that might be held by stakeholders is that program participants (of a health education program, for example) learn better by doing rather than by listening to someone tell them how to do something (assuming, of course, that this is not yet empirically proven). Paradigmatic assumptions are also referred to as descriptive assumptions (Zaman and Rabbi 2010) and value assumptions (Cowgill 1964; Niemeyer 1955). Sociologists have used the term "value" in at least four basically different ways: (1) as universal absolutes, (2) as attributes of objects, (3) as internal states of man, and (4) as ways of behaving toward things. Something has or is of value if and when people behave toward it so as to retain or increase their possession of it (Cowgill 1964). Value assumptions create the perspectives in which we see reality and the object of research. If one assumes that a policy or approach is good, one asks certain questions and marshals facts in a certain way (Niemeyer 1955).

Forms of Paradigmatic Assumptions

Ontological Assumptions

These are often used in reference to research paradigms and they are assumptions we make about the nature of reality. For example, one evaluator might examine the well-being of a community with the assumption that there is no such a thing as a definite state of well-being in the community; and that this community's well-being has as many shapes as the beliefs that community members have about their well-being. On the other hand, an evaluator may premise his study of a community's well-being on the ontological assumption that the state of well-being is an objective truth that can be clearly ascertained. Ontological assumptions are discussed in further detail in Chap. 7.

Epistemological Assumptions

These are assumptions about how reality can be ascertained. One's assumptions about the nature of reality (ontology) are related with assumptions on how reality

can be studied (epistemology). As in the example on ontological assumptions above, the evaluator who assumes that the state of a community's well-being is an objective fact also assumes that this "objective fact" can be studied without bias. On the other hand, an evaluator who discards the "objective reality" assumption inevitably assumes that it is impossible to study a community's well-being without bias. The discussion on epistemological assumptions is continued in Chap. 7.

Axiological Assumptions

Axiology describes what is valued or considered good (Mingers 2003). Axiological assumptions have to do with whether and how we expect values to meddle in the processes or the products of inquiry. Axiology is related to epistemology. The assumption of objective reality may imply that values either of the evaluator or the study participants; confound the process and product of inquiry. On the other hand, the assumption of subjective reality implies that values enrich rather than confound the process and product of inquiry.

Cultural Assumptions

These emerge from cultural influences individuals raised or who live in different parts of the world are likely to have different "paradigms" or "world views" (Kies 1995). I.e. cultural bias or the lens of culture through which one interprets the world.

Idiosyncratic Assumptions

These are the assumptions that result from events—positive or traumatic—in individuals' personal experiences (Kies 1995). Some assumptions may have supporting evidence; these are called warranted assumptions. However, many assumptions in this category have little or no supporting evidence at all (and are called unwarranted assumptions).

Prescriptive Assumptions

Prescriptive assumptions, as defined by Brookfield, are assumptions about what should happen in certain situations. For example, if stakeholders hold the paradigmatic assumption that people learn better by doing, they might also hold the prescriptive assumption that good community educators should engage participants in learner- centered and participatory activities.

Categories of Prescriptive Assumptions Include

Legal Presumptions

Legal assumptions apply in litigation. A legal presumption is an initial probabilistic estimation by a court (in absence of evidence) of some facts. According to Bernardo et al. (2000) a presumption is either conclusive or rebuttable. It also can describe a judicially applied prediction of factual or legal probability (Legal Usage in Bernardo et al. 2000, p. 1). These are different from evidentiary burdens of proof, which refer to whether or how the parties can extricate the proposition at hand through the presentation of evidence. The legal presumption can be divided into two types of situations: (1) the burden of production focuses on which litigant has the duty to come forward and present the evidence (or risk losing); (2) the burden of persuasion refers to applicable criteria, the court would use to update its initial finding on the presumption of evidence (e.g., preponderance, clear and convincing, beyond reasonable doubt, etc.) (Bernardo et al. 2000).

Metaphoric Assumptions

These are assumptions implied in analogical expressions. For example, in metaphors such as "I'm on top of the world" and "I'm feeling down today," it is assumed that being "up" and being "upright" are associated with good, while being "down" may be associated with negative or bad (Kies 1995).

Intellectual Assumptions

An intellectual assumption is a bias toward rationality (Kies 1995). It is a normative belief that actions that are not backed by facts are not justified.

Causal Assumptions

Causal assumptions, according to Brookfield, are assumptions about how different parts of the world work and about the conditions under which these can be changed. They are also known as predictive assumptions (Eliot 1937). They have to do with expressions of attitude and opinion which have a predictive character, or represent a readiness to act on the assumption of some probability. They can be hypotheses, prognoses, reputations, stereotypes, and epithets (Eliot 1937). Assuming that, the use of participatory learner-centered education would significantly improve knowledge and health behavior is an example of a causal assumption.

Categories of Assumptions by Degree of Articulation

Assumptions can also be categorized on basis of how much they have been or not been explicated. The explication of assumptions is the core of this text. Degrees of explication are discussed at greater depth in Chap. 7. In this section, assumptions are categorized as explicit or implicit.

Explicit Assumptions

These are assumptions of whatever category prescriptive, descriptive, or paradigmatic, that are formally acknowledged and documented.

Implicit Assumptions

Implicit assumptions are those that have not been articulated or verbalized. They will be explored in greater detail in this section. Implicit assumptions can be categorized into conclusions, unstated premises, and presuppositions and rationalizers (Ennis 1982).

Conclusions

This is a proposition that is not fully established. For example, to see someone wearing a cap and assumed that they are going out (Ennis 1982).

Unstated Premises

A premise or presupposition is a proposition with in an argument; the strength of the argument depends on the validity of a premise (Ennis 1982). For example, the conclusion (or argument) that one is going out because he has put on a cap may be based on a premise that people do not put on caps until they have to go out of the house. If this is not necessarily the case, then the conclusion is undermined.

Rationalizers

These assumptions that are consciously or unconsciously employed to support a conclusion, to back the argument, or to rationalize a position. Sometimes they are unstated reasons that the individual uses consciously or subconsciously as a basis for argument or action. These commonly include bias that are sometimes contradictory to our conscious beliefs; they may generate social stereotyping, prejudice, and sometime discrimination (Banaji and Greenwald 1994; Blair 2002;

Blair et al. 2001; Greenwald et al. 2003). Stereotypes are a common example of needed and used implicit assumptions.

Stereotypes are beliefs about the attributes of social groups (Banaji and Hardin 1996). "Stereotyping, like other cognitive processes, consists of both automatic and controlled components, and the particular form of automaticity that is involved (e.g., awareness, intentionality, efficiency, and controllability). Such beliefs operate unconsciously" (Banaji and Hardin 1996, p. 136). Thus the individual's attitudes and actions may be unintentional, unaware, implicit, automatic, and involuntary. Research shows that every one of us brings a lifetime of experience and cultural history that shapes how we understand and interact with the world around us. "… as we become aware of our hypotheses, we replace our belief in a just world with a view of the world in which bias plays a role. Since this is a state of affairs we wish were otherwise, we prefer not to acknowledge it. But we can learn" (Valian 1999, p. 3003). A common example of stereotyping is the gender schemas.

Gender schemas are nonconscious hypotheses that people hold about the different characteristics of males and females. For example, one schema holds that females are nurturing, communal, and do things out of concern for other people. Another holds that males are more rational and capable of acting independently. While schemas may not predict or match with actual behavior, they influence a great deal of how people view the world around them. For example, research by Valian (1999, 2004) shows that for some people, the schema that people have for being a professional person overlaps much more with the schema for being male than it does with the schema for being female.

Research by Heilman et al. (2004) shows that when women are acknowledged to have been successful, they are less liked and more personally derogated than equally successful men. These negative reactions occur only when the success is in an arena that is seen as distinctly male in character. Being disliked by colleagues, whether for objectively valid reasons or because of some perceived schema, can have career affecting outcomes, both for overall evaluation and for recommendations concerning organizational reward allocation. These results were taken to support the idea that gender stereotypes can prompt negative bias in evaluative judgments of women even when these women have proved themselves to be successful and demonstrated their competence.

Also, a study by Trix and Psenka (2003) examined over 300 letters of recommendation for medical faculty positions at a large American medical school in the mid 1990s. They found that letters of recommendation written for female applicants differed systematically from those written for male applicants in the extremes of length, in the percentages lacking in basic features, in the percentages with doubt raisers and in frequency of mention of status terms. Use of possessive phrases reinforced gender schema that tend to portray women as teachers and students, and men as researchers and professionals.

Countless studies have provided evidence of such gender schemas (e.g. Conley 1998; Fried et al. 1996; Stokoe and Smithson 2001; Sean et al. 1996; Buchner and Wippich 1996; Good et al. 2008). Numerous studies show evidence of

stereotyping based on other social characteristics. In Akalis et al. (2008) study where participants visualized four scenarios involving black or white would be criminals, participants demonstrated significantly more negative implicit stereotypes toward blacks as a group compared with a condition in which white would be criminals were more salient in memory. In Shutts et al. (2010)'s study, 3-year-old children were asked to choose between objects or activities that were endorsed by unfamiliar people who differed in gender, race (white, black), or age (child, adult). Children demonstrated strong preferences for objects and activities endorsed by children of their own gender, but less consistent preferences for objects and activities endorsed by children of their own race. Children selected objects and activities favored by people of their own gender and age. Children did not acknowledge the influence of these social categories (See also Davies et al. 2005; Rydell et al. 2009; Steele et al. 2002).

The exploration of implicit assumptions above indicates that they have a great influence on how individuals comprehend and interact with the world around them. What evaluators have to learn from this is that one cannot sufficiently understand development programs if the assumptions underlying such programs remain implicit. It is, thus, necessary for stakeholders of complex development programs to examine their underlying assumptions—surfacing and exploring the implications for their work—because thoughts and feelings outside of conscious awareness or conscious control (which may well be different from our explicitly endorsed beliefs and values) can influence perception, judgment, and action. Explicating these assumptions and/or beliefs makes it possible to examine them and makes it easier to evaluate programs that have been designed based on these assumptions.

References

Akalis, S. A., Banaji, M. R., & Kosslyn, S. M. (2008). Crime alert! How thinking about a single suspect automatically shifts stereotypes toward an entire group du bois review. *Social Science Research on Race, 5*(2), 217–233.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7*(3), 136–141.

Banaji, M. R., & Greenwald, A. G. (1994). Implicit stereotyping and prejudice. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 17, pp. 55–76). Hillsdale, NJ: Erlbaum.

Bennett, G. E. (1933). Assumptions. *The Accounting Review, 8*(2), 157–159.

Bernardo, A. E., Talley, E., & Welch, I. (2000). A theory of legal presumptions. *Economics and Law, 16*(1), 1–49.

Brookfield, S. D. (1995). *Becoming a critically reflective teacher*. San Francisco, CA: Jossey-Bass.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242–261.

Blair, I. V., Jennifer, E. M., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*(5), 828–841.

Buchner, A., & Wippich, W. (1996). Unconscious gender bias in fame judgments? *Consciousness and Cognition, 5*, 197–220.

Cowgill, D. O. (1964). Value assumptions in recent research on migration. *International Migration Digest, 1*(1), 29–33.

Conley, F. K. (1998). *Walking out on the boys*. New York: Farrar, Straus and Giroux.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88*(2), 276–287.

Eliot, T. D. (1937). Reactions to predictive assumptions. *American Sociological Review, 2*(4), 508–517.

Ennis, R. H. (1982). Identifying implicit assumptions. *Synthese, 51*(1), 61–86.

Fine, E., & Handelsman, J. (2006). *Reviewing applicants: Research on bias and assumptions*. Madison, WI: University of Wisconsin.

Fogarty, R. J. (1994). *How to teach metacognitive reflection*. New York: Corwin.

Fried, L. P., Francomano, C. A., MacDonald, S. M., Wagner, E. M., Stokes, E. J., Carbone, K. M., et al. (1996). Career development for women in academic medicine: Multiple interventions in a department of medicine. *Journal of the American Medical Association, 276*, 898–905.

Harste, J. C., Woodward, V. A., & Burke, C. L. (1984). Examining our assumptions: A transactional view of literacy and learning. *Research in the Teaching of English, 18*(1), 84–108.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197–216.

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*(1), 17–28.

Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology, 89*(3), 416–427.

Kanazawa, S. (1988). In defense of unrealistic assumptions. *Sociological Theory, 16*(2), 193–204.

Kies, D. (1995). Assumptions. Retrieved December 20, 2011 from http://papyr.com/hypertextbooks/comp2/assume.htm.

Kotze, D. A. (1987). Contradictions and assumptions in community development. *Community Development Journal, 22*(1), 31–36.

Mingers, J. A. (2003). Classification of the philosophical assumptions of management science. *Journal of the Operational Research Society, 54*(6), 559–570.

Niemeyer, G. (1955). A query about assumptions on international organization. *World Politics, 7*(2), 337–347.

Read, S. (2003). Freeing assumptions from the Liar paradox. *Analysis, 63*(2), 162–166.

Russell, B. (1904). Meinong's theory of complexes and assumptions. *Mind, New Series, 13*(51), 336–354.

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology, 96*(5), 949.

Sean, C., Draine, S. C., & Greenwald, A. G. (1996). Modeling unconscious gender bias in fame judgments: Finding the proper branch of the correct (multinomial) tree. *Consciousness and Cognition, 5*, 221–225.

Sherman, B., & Harman, G. (2011). Knowledge and assumptions. *Philosophical Studies, 156*(2011), 131–140.

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science, 13*(4), 599–610.

Sosa, E. (1974). How do you know? *American Philosophical Quarterly, 11*(2), 113–122.

Steele, C. M., Spencer, S. J., & Aronson, J. (Eds.). (2002). *Contending with group image: The psychology of stereotype and social identity threat (Vol. 34, pp. 379–440)*. San Diego: Academic Press, Inc.

Stokoe, E. H., & Smithson, J. (2001). Making gender relevant: Conversation analysis and gender categories in interaction. *Discourse & Society, 12*(2), 217–244.

Trachman, M., & Bluestone, C. (2005). What's love got to do with it? *College Teaching, 53*(4), 131–136.

Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society, 14*(2), 191–220.

Ullman-Margalit, E. (1983). On presumption. *The Journal of Philosophy, 80*(3), 143–163.

Valian, V. (1999). *Why so slow? The advancement of women*. Cambridge: MIT Press.

Valian, V. (2004). Beyond gender schemas: Improving the advancement of women in academia. *Hypatia, 20*(3), 198–213.

Zaman, M. H. & Rabbi, R. E. (2010). What are the descriptive assumptions in critical reading? Retrieved December 12, 2011 from http://criticalreadingsblogspot.com.

Chapter 7
Why are Assumptions Important?

Abstract Assumptions are the foci for any theory and thus any paradigm. It is also important that assumptions are made explicit, and that the number of assumptions is sufficient to describe the phenomenon at hand. Explication of assumptions is even more crucial in research methods used to test the theories. As Mitroff and Bonoma (Evaluation Quarterly 2(2):235–260, 1978) have eloquently put it: "the power of an experiment is only as strong as the clarity of the basic assumptions which underlie it. Such assumptions not only underlie laboratory experimentation but social evaluation research as well." A scale of articulation of assumptions is represented: (a) Very ambiguously tacit assumptions held in divergent beliefs; (b) Tacit but more obvious assumptions-where parties more or less have shared although unexpressed perceptions or beliefs e.g., legal assumptions in litigation; (c) Informally, explicit assumptions e.g., indirectly expressed or inherent in shared stories, norms, un-codified symbols, and myths; (d) Assumptions that are made explicit. E.g. in theories and models, policies and programs, research and methodology but untested; and (e) Explicit and tested assumptions.

> *Apparently, assumption is believed to provide an easy or lazy person's way out of what seems to be, at least at first glance, a perplexing situation. At least, one is not justified in going ahead and making some assumption which moderately expressed is wholly an arbitrary one. Accuracy of interpretation never can be supplanted by assumption. I assume one thing, you assume something else, and the other fellows each assume something different from each of the others. Grant the acceptance of assumption, and then any one of a 100 different solutions is correct. Could anything be more ridiculous?*
>
> Bennet 1933, p. 158

Introduction

Three functions for assumptions are discussed in this chapter:

(a) Guiding argumentation, evidence generation, and conclusions
(b) Guiding the selection of research methodology
(c) Providing the foci for paradigms and theories

In the first two functions, argumentation and selection of methodology, assumptions are often used consciously or even explicitly. It is in the third function providing foci for paradigms and theories that assumptions often perform a latent and yet the most powerful role.

Assumptions Guide Argumentation, Evidence Generation, and Conclusions

One reason, why assumptions are important, is that they frame and guide,directly or indirectly,the processes of argumentation, evidence generation, and conclusions. This discussion draws mainly from the use of assumptions in litigation. In the legal system, a presumption implies a rule of law that courts and judges shall draw a particular inference from a particular fact, or from particular evidence, unless and until the truth of such inference is disproved (Ullman-Margalit 1983). A presumption of fact is when another fact is assumed from established basic facts. A presumption of law is when a rule of law requires an assumption to be made (Ullman-Margalit 1983). Presumption is a rational prerequisite for arrival at a variety of decisions. For example, legal presumptions are used in the justice system to arrive at verdicts. Some examples of legal presumptions are offered by Ullman-Margalit 1983, p. 144: "…that a person who, without reasonable explanation, has not been heard from for at least seven years is dead; that a marriage regularly solemnized is valid; that if A buys property from B and directs it to C, C was intended to be a trustee for A; that when in a common disaster the death occurs of two or more persons related as to inheritance, the younger (or healthier) person survived longer. Also, that a person accused of crime is innocent; that every person is sane; that a person intends the natural consequences of his or her actions."

In situations where actions have to be decided upon in the light of insufficient information and often under external pressures and constraints, presumption rules are used. The jury might follow a presumption rule in generating a conclusion. For example, the jury may use the presumption of death upon the expiration of 7 years of unexplained absence, in examining whether there are explanatory facts to account for the absence. If such facts are absent, death is concluded as a fact. The question of what it takes to bring about the refutation of a presumption concerns the twin notions of the burden of proof and the benefit of the doubt. For example, if

and as long as there is no (sufficient) evidence, that the absent person did not die, the jury is likely to conclude that the absent person is dead. The presumption is discarded only on basis of evidence. The evidence has to be produced by the presumption proponent and it is up to the other party, the presumption opponent, to discredit this evidence and discard the presumption.

It is important to note that the generation of evidence is a central part of the "litigation" process. Presumption, in this case, plays a role of focusing parties toward or away from a line of argumentation in the presentation of evidence; and more so, the presumption presents an obvious line of conclusion to confirm or discard the presumption (Mir and Watson 2001).

While presumption has been sometimes defined as a form of tacit assumption, the discussion of legal presumptions in the preceding section shows that such presumptions are not tacit, since all parties usually share the same understanding of the presumptions, even when they are not specifically stated. This kind of clarity is uncommon with the "underlying assumptions" in program theory. Undoubtedly, explicating "underlying assumptions" can have a key role in clarifying which questions should be the focus of evaluation and how fairly precise and relevant conclusions can be stated from data.

Assumptions in the Selection of Methodology

> Looking for assumptions in research is something like looking for shadows in the late afternoon. They are everywhere. And once one starts noticing them he suddenly becomes aware that the world is full of them. In fact, if one contemplates the assumptions that underlie nearly everything he does, he quickly decides he had better stop such contemplating lest he conclude that nothing at all can be done with sufficient confidence to warrant doing it (Scates 1940, p. 241).

Another key role of assumptions, in program evaluation and generally in social science research, is the selection of methodology. Choices throughout the research process are always "rendered from within a complex constellation of philosophical assumptions" (Green et al., 2007, p. 111) about the nature of reality (ontology), how we get to know what we know (epistemology), the inclusion of values (axiology), and the nature in which their research emerges (methodology) (Campbell and Wasco 2000; Morgan and Smircich, 1980; Cook and Fonow 1986; Guba and Lincoln 1989; Hofer and Pintrich 2002; Granello 2002, Wegerif 2008).

One of the first decisions an evaluator has to make in designing an evaluation study is the choice of a philosophical or paradigmatic position in which the study is anchored (Hawkins 1978). Each philosophical or paradigmatic position is defined by a set of "basal" or "meta theoretical" (Hawkins 1978, p. 142) assumptions (Creswell et al. 2007; Brannen 2012; Schwandt 2001). These assumptions (referred to as paradigmatic assumptions in this text) directly influence methodology, methods of data collection, the evaluator—participant relationship, measures of

quality, and how evaluation findings should be reported (Carter and Little 2007; Becker and Niehaves 2007).

Ontological Assumptions

Ontology relates to the essence or the nature of being (Hougon Muffin Company 2004, p. 972). It is the analysis of "what is" and "how it is" (Becker and Niehaves 2007). Ontology is the "reality" that researchers investigate (Healy and Perry 2000; Scott and Usher 1996). Implicitly or explicitly, a researcher has to accept or reject the notion that there is a single, objective, real world (Robson 1993). This is the primary ontological decision point for researchers (Campbell and Wasco 2000). If a researcher accepts the premise of objective reality, then the goal of science is to discover the structure and function of that singular world. If a researcher does not accept this premise, then the goal of inquiry is to understand how we construct and interpret our realities (Campbell and Wasco 2000; Mir and Watson 2001).

Epistemological Assumptions

Epistemology concerns its self with "analyzing the way human beings comprehend knowledge about what is perceived to exist" (Becker and Niehaves 2007, p. 201). Epistemological assumptions are assumptions about the nature of knowledge: "its presumptions and foundations, and its extent and validity" (Hougon Muffin Company 2004, p. 472; also see Campbell and Wasco 2000; Morgan and Smircich 1980; Cook and Fonow 1986; Guba and Lincoln 1989; Hofer and Pintrich 2002; Granello 2002; Wegerif 2008, etc.).

Epistemology is intertwined with ontology, since claims about what exists in the world (reality) imply claims about how that existing reality can be known (Scott and Usher 1996). For example, if a researcher accepts the notion of an objective reality, then the researcher inexorably assumes "a position of objective detachment, free from bias so as to be able to capture that reality accurately" (Campbell and Wasco 2000, p. 779). On the other hand, if a researcher does not accept the ontological notion of objective reality, there would not be need to assume a position of objective detachment (Mir and Watson 2001).

Scott and Usher (1996) describe the evolution of epistemology from various debates about how reality can be known. Historically, the traditional source of knowledge was belief and opinion. With the advent of enlightenment, this source of knowledge was considered invalid and thus dismantled. The more appropriate approach was to "ground the validity of knowledge in scientific method in the form of measurement, testability and the right use of reason" (Scott and Usher

1996, p. 11). Thus, observation and experimentation (empiricism) became the accepted route to acquisition of valid knowledge.

Guba and Lincoln (1989) in Campbell and Wasco (2000, p. 779) identify four primary epistemological theories that clarify the nature of knowledge: positivism, realism, critical theory, and constructivism. Healy and Perry (2000) summarize these four paradigms into three categories: The first category is positivism, and it lies at the objectivism end of the objective–subjective continuum. The second category is related to critical theory and constructivism, and lies on the subjectivism end of the objective–subject continuum. The third category, realism, seems to lie in the middle of these two.

Positivism

Positivism is premised on the ontological assumption of an objective reality. Therefore, the goal of inquiry is to explain, predict, and ultimately control that reality. Hypotheses about objective reality are generated, tested, and verified, primarily though experimental methods. Throughout the process of research, the researcher must engage in a distant, objective stance to remain free from biases that could interfere with obtaining knowledge. To the positivist, the identity of the "knower" (i.e., inquirer), is not especially as relevant, as proper use of the scientific method, by which anyone, should capture objective reality (Hofer and Pintrich 2002; Campbell and Wasco 2000). The researcher's personal beliefs and values could interfere with assessment of that objective reality; therefore, caution has to be taken to ensure that these human factors do not enter into the scientific process (Perry 1998; Clark 1998; Healy and Perry 2000; Shah and Corley 2006; Gioia and Pitre 1990).

Realism, like positivism, is based on the notion that an objective reality exists; that there is a real world that researchers could discover. However, realism differs from positivism by noting that it is probably impossible to capture reality in a pure, unsullied, and unbiased form (Healy and Perry 2000; Shah and Corley 2006; Campbell and Wasco 2000). Thus, as humans, researchers are naturally flawed and incapable of capturing the pure, real, social, and physical world with absolute accuracy (Campbell and Wasco 2000).

Critical theory posits that reality is interpreted through social, political, cultural, economic, ethnic, and gender values, and therefore there is no one objective reality (Campbell and Wasco 2000; Healy and Perry 2000). What is considered to be knowledge is not "pure fact", because it is filtered through these various lenses. The researcher's values, beliefs, and experiences influence how research questions are formed, data are collected, and findings are interpreted. "Because it is not possible to remove or partial out these factors, researchers must articulate how their individual experiences shape their research" (Campbell and Wasco 2000, p. 780).

Critical science research emphasizes the need for a researcher's awareness of social values and beliefs during inquiry. This awareness provides a way of thinking

critically about the world, including the interconnected networks which provide meanings to the concepts and values held (Borg and Gall 1983; Bredo and Feinberg 1982; Healy and Perry 2000; Coomer 1984; Shah and Corley 2006).

In discarding the notion of objective reality, constructivism takes critical theory a step further by arguing that reality is a social construct. Thus, truth is a particular belief system held in a particular context (Healy and Perry 2000). There are as many truths or realities as there are visions of reality. People's visions of their realities are shaped by "social factors, such as gender, race, class, culture, and economics" which "are not merely lenses through which we see reality" but rather "agents shaping how we construct our visions of what constitutes our individual realities" (Campbell and Wasco 2000, p. 780).

The above discussion on paradigms and their ontological assumptions is not exhaustive. It highlights the main positions along the objective–subjective continuum. While different methods used by evaluators are in one way or the other premised on paradigmatic assumptions, it is not clear if evaluators recognize them enough in their day-to-day methodological decisions. In fact, some researchers think of such discussions as cryptic and less practical (Carter and Little 2007). Some have argued that research methods cannot be perfectly aligned with paradigms. Evaluation methodology is about more principles, standards, or "things that work well or not" that might be too messy to group in discrete paradigms (Hammersley 1995; Bryman 1988; Bhaskar 1975; Bonell 1999; Sale et al. 2002). For example, Bonell (1999) argues that the equation of quantitative methods with the positivist paradigm is "too simplistic" (Bonell 1999, p. 21); that quantitative and experimental methods should not be "equated to the positivist paradigm, if such a thing can even be said to exist" (Bonell 1999, p. 21).

However, a choice of methodology that does not take paradigmatic assumptions into account is not defensible (Greene and Curucelli 1997; Yauch and Steudel 2003). Whether they are recognized or not, paradigmatic assumptions have implications on the nature of methods, processes, and conclusions an evaluation is expected to produce. Consider an evaluator, Jane, who in evaluating a maternal and child health program in a community intends to examine why mothers are reluctant to deliver at the health facility (which is a safer option); but choose to deliver at home. From a more constructivist perspective, Jane will be generating knowledge on the consumption of antenatal services in collaboration with the mothers she is studying; and while her perspectives as an evaluator are important in shaping the study, she takes it that valid conclusions will not be made without understanding demand for antenatal services from the mothers' point of view. Also, the creation of this knowledge will largely depend on the nature of interaction and relationships she will have with participants, for example, relationships of trust versus suspicion. The kind of findings she is likely to generate cannot be the same as those she would generate in another community at a different time even on the same evaluation question, because the nature of relationships cannot be exactly the same.

As a constructivist, Jane will be involved in the study entirely and actively, without withholding her values and subjectivities from it. For example, she might

have delivered three children herself in a better or worse setting; she might have prejudices against certain health care personnel. Thus, she would have to document her own subjectivities such that readers of her evaluation can judge how these subjectivities could affect the validity of her conclusions.

From a constructivist perspective, Jane does not need to start with definite measures of mothers' attitudes and perceptions or beliefs about delivering at the health center. She may begin her study with broad questions, such as "would a mother want to return to the health facility for another delivery; and why? Or what's the quality of midwife-client relationships?" Her interactions with participants through interviews and observations will dynamically generate an array of themes, insights, and definitions e.g., of midwife-client relationship and if or how these could constitute barriers to seeking maternity services. If distance from the health facility emerges as a barrier, she may be more concerned about understanding if mothers think the facility is far or inaccessible than the exact distance between mothers' homes and the facility.

On the other hand, from a positivist perspective, Jane would be attempting to understand the reality of health seeking among mothers and with the intent to generalize these findings as much as possible. She tries to measure mothers' real attitudes and beliefs about antenatal services delivered at the community health center. She will, thus, try to use methods of data collection that limit the extent of bias in the evaluation, either through selection of participating mothers, or in collecting and processing the data. Thus, her conclusions can be inferred to similar communities.

For example, her experiences as a mother and related prejudices should not meddle in her inquiry. She does not have to doubt responses dissimilar to her experiences or necessarily affirm those that she is familiar with. She needs to start her study with clear definition of constructs, e.g., outlining the attributes of a good or bad midwife-client relationship and if these exist. She should have read about a range of barriers to seeking antenatal services and if distance was one of them, she would need to decide exactly how far is far and how near is near in terms of miles or kilometers.

The above example might have exaggerated some of the differences in paradigmatic positions in the effort to show that implications of paradigmatic assumptions in selection of research methodology need not to be ignored. This is especially true when different methods, based on different paradigmatic assumptions, have to be triangulated within an evaluation. It is argued that the quality of mixed methods evaluation would depend on how evaluators take into account the paradigmatic assumptions of methods (qualitative and quantitative) that they triangulate within a study (e.g. Green and Curucelli 1997). Others have argued that this "requirement" might be one of the challenges of mixed methods research. In a study on barriers to integrating qualitative and quantitative data, Bryman (2006) interviewed 20 mixed methods researchers and he learned that even when researchers used both quantitative and qualitative designs, they often did not integrate the two forms of data. One of the barriers to integrating qualitative and quantitative data within mixed methods studies was that the two designs are

underpinned by different paradigmatic assumptions. Bryan refers to this barrier as that of "marrying an objectivist account with a constructivist one based on people's discursive accounts" (Bryman 2006, p. 16); e.g., the constructivist account denies "the reality of the very phenomena that the objectivist account seeks to investigate" (Bryman 2006, p. 16); For example, in a study on poverty, the qualitative design might regard respondents' accounts from a constructivist standpoint that is ambivalent of whether the people are really poor-or it is just what they think of their well-being or not; yet, the quantitative design tries to furnish data that provides a more or less indisputable evidence of poverty.

There are different views on the extent to which paradigmatic assumptions should influence methods selection in mixed methods research (Greene and Curucelli 1997)-

- The purist view that the paradigms are underpinned by different and conflicting paradigmatic assumptions and that it is unreasonable to combine these assumptions in a way that allows mix of the methods on which they are based.
- The dialectical view that goes beyond the purist view to suggest that the differences among paradigms and their paradigmatic theoretical assumptions do not only exist, but they are important and irreconcilable. These differences need to be honored in ways that maintain the integrity of the different paradigms (Greene and Curucelli 1997).
- The pragmatic view that in the real world, the paradigms describe but do not dictate how research should be done; that methods can be mixed without necessarily mixing the paradigms or paradigmatic assumptions on which they are based.

Some evaluators may not explain if or not their use of mixed methods was influenced by any one of the above views. When this is not explained in the evaluation reports or articles published from such evaluation, it might imply that they either did not consider the significance of paradigmatic assumptions in their mixed methods research or they may have taken a pragmatic view of the paradigmatic assumptions. Even when evaluators have taken a more flexible view of these assumptions, it is still necessary to describe them. As Patton, one of the proponents of the pragmatic view argues, the purpose of describing alternative research paradigms is to sensitize researchers and evaluators to the ways in which their methodological prejudices, derived from their disciplinary socialization experiences, may reduce their methodological flexibility and adaptability. The purpose of describing how paradigms operate is to free evaluators from the bonds of allegiance to a single paradigm" (Patton 2011, cited in Greene and Curucelli 1997, p. 9).

In defense of the paradigmatic assumptions, Greene and Curucelli (1997, p. 11) argue that "the pragmatists, especially the extremists in the camp, may err in granting too little authority to paradigms as shaping practical inquiry decisions. While the demands of the inquiry context are important, they may be insufficient guides for evaluative inquiry." Greene and Curucelli argue for a balanced approach to the application of paradigmatic assumptions. "Extremists with in the dialectical stance may also err in granting too little authority to the inquiry context

as shaping practical inquiry decisions. The middle position here consists of a balanced, reciprocal relationship between philosophy and methodology, between paradigms and practice. This position rejects extremism…, honors both the integrity of the paradigm construct and the legitimacy of contextual demands and seeks a respectful, dialogical interaction between the two in guiding and shaping evaluation decisions in the field." In support of this balanced approach, Gioia and Pitre (1990) have argued that paradigm boundaries are permeable. And they give an account of how to encourage a more positive proliferation across paradigms and argue that "Although the central assumptions of the paradigms clearly are at odds the boundaries between them tend to be ill-defined and 'blurred'" (Gioia and Pitre 1990, p. 591). Since the paradigmatic dimensions of subjectivity/objectivity and stability/change are continua, it is difficult to divide them neatly. Hence it better to conceive them as transition zones.

Assumptions are the Foci for Paradigms and Theories

Tacit assumptions can be traced at the core of paradigms. A paradigm is a world view built on mostly implicit assumptions (Patton 2011; Mingers 2003). Paradigms tell us what is important, legitimate, and reasonable. Patton (2011) argues that the power of paradigms is also their weakness. The implicit assumptions on which they are built make action possible, yet those assumptions are rarely questioned. In his book *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies-How We Construct Beliefs and Reinforce Them as Truths*, Micheal Shermer (2011) argues that people form such belief from a variety of subjective, personal, emotional, and psychological reasons in the context of environments created by family, friends, colleagues, culture, and society at large. After forming our beliefs we then defend, justify, and rationalize them with intellectual reasons, arguments, and explanations. According to Shermer, beliefs come first, and explanations for beliefs follow.

In reference to research, "paradigm" has been used to refer to a combination of basic assumptions about the nature of science and society (Shrivastava and Mitroff 1984; Mingers 2003). It is a fundamental core of assumptions that form the basis for other assumptions and jointly underlie all human inquiry-scientific and non-scientific, formal, and informal. Paradigms guide the conduct of inquiry by defining legitimate research questions in a field, describing appropriate methods for addressing these questions, outlining the broad scope of an inquiry, and encapsulating a researcher's assumptions in inquiry.

Paradigms are characterized by a number of features including (a) cognitive elements; (b) cognitive operators; (c) reality tests; (d) cognitive maps of the domain of inquiry; (e) degrees of articulation; and (f) metaphors (Shrivastava and Mitroff 1984). All these components reflect in themselves various forms of tacit

and explicit assumptions that influence how individuals interact with the world around them as well as the conduct of inquiry.

Cognitive Elements

A cognitive element is an orientation toward the belief in the actual or potential existence of things (events, relationships, etc.), whether experiential or supernatural (Shrivastava and Mitroff 1984; Ingwersen 1996; Warner 1978). For example, beliefs about the existence or nonexistence of supernatural powers is a cognitive element (Warner 1978). "The individual who believes something tends not to question the factual accuracy or empirical verifiability of what they believe." Cognitive elements include, among other things, cognitive categories, and bits of data that are taken for granted or regarded as so basic that they are beyond doubt. These bits of data are fundamental units of information that support a person's inquiring system or concept of the world (Kerem et al. 2001).

Cognitive Operators

Cognitive operators are the methods by which individuals process information and generate meaning out of large amounts of data (John and Kieras 1996). Cognitive operators include classification schemes, models, analytical devices, and commonsense theories with which individuals approach inquiry (Shrivastava and Mitroff 1984).

Reality Tests

Reality tests are a key component of a paradigm. It is reality tests that enable individuals to test or validate cognitive elements, cognitive operators, and knowledge or information itself (Lynn et al. 1989). Thus, the information individuals acquire from their environment undergoes continuous corroboration with their experiences. An individual's grounding in reality is evaluated in view of their experiences over time (Shrivastava and Mitroff 1984).

Cognitive Map of the Domain of Inquiry

This element of a paradigm refers to a researcher's assumptions about the scope of inquiry and the nature of its boundaries. This scope of a study could be based on a discipline, subject matter, or research paradigm (Crossan et al. 1995).

Degree of Articulation

This refers to the extent to which the assumption(s) embodied in the other four elements have been articulated and codified. It also reflects the degree to which the individual's paradigm or frame of reference can be shared by others. Making something explicit does not necessarily mean the understanding is shared (Crossan et al. 1999).

Metaphors

Metaphors are a medium that can be used by individuals to exchange information on their paradigms or assumptions (Crossan et al. 1999). Individuals use metaphors to help explain their intuition to themselves and to share it with others (Tsoukas 1991).

Assumptions in Theory

Theory is another place we find assumptions. The connection between paradigms and theories is described by Kuhn (1970), White and McCain (1998), and Hjørland (2002), among others, by arguing that "paradigms", are made up of formal theories, experiments, and methods. Paradigms are a source of theories. The scope of a paradigm is redefined by testing and refining derived theories, and developing new methods.

Babbie (2007, p. 43) defines theories as "systematic sets of interrelated statements intended to explain some aspect of social life." Theory can also be defined as a set of statements, organized in a characteristic way, and designed to serve as partial premises for explaining, as well as predicting, an indeterminately large (and usually varied) class of economic phenomena (Nagel 1963). Theories flesh out paradigms, whereas a paradigm provides a way of looking and is grounded in assumptions about the nature of life, a theory explains what we see. Whereas a law is an observed regularity, a theory is a systematic explanation of observations that relate to a particular aspect of life (Babbie 2007). Theories explain observations by means of concepts, which are abstract statements representing classes of phenomena with in the field of study. A variable is a concept that contains a set of attributes.

Since theories, as Babbie suggests, make statements about relationships between concepts, they make certain assumptions about the conditions under which such relationships would exist. Sometimes the very relationships among concepts that are stated in theories are made in form of assumptions or hypotheses. It can be said that assumptions are the glue that hold theories (concepts, the building blocks of theories) together. Hawkins (1978, p. 141) supports this

argument: "Most social theories, contain many socio philosophical assumptions about human nature… When policies geared toward promoting social change are proposed, an investigation of the adequacy of these assumptions is essential. Any proposal for directed social change assumes that human beings will act or react in predictable ways to a given stimulus for change."

Kanazawa (1988) argues that assumptions and hypotheses are the foci that make up theory. Hypotheses are framed from assumptions. A theory is tested by subjecting its hypotheses (framed from assumptions) to empirical evidence.

Assumptions Operate in a Theory in Two Ways: Explicitly and Implicitly

(a) Assumptions expressed as hypotheses and propositions that make up a theory: For example, a theory to bring about positive behavior among youth may be premised on the assumption that youth would more easily acquire a new behavior if they observe it practiced by someone they admire. In this case, assumptions are explicit foci of theory.

(b) Assumptions may exist implicitly in a theory. The functioning of assumptions within theory depends on what and how a given theory is attempting to describe or predict. Kanazawa (1988) argues that the empirical world is highly complex, with many variables operating. It follows that it takes a large number of statements to describe some part of the empirical world accurately. The empirical complexity in its entirety is not reducible to an arbitrarily small set of descriptive statements. For example, if it takes a given number of statements (say ten) to describe a phenomenon adequately, using two statements would ignore 80 % of its complexity. One cannot describe more of the empirical complexity with fewer statements (Kanazawa 1988). So the main question becomes, "What happens to that 80 % of a complexity about which a theory is silent?" It is in essence assumed. Whether it is assumed existent or nonexistent, it is assumed implicitly.

Looking at youth behavior change theory; what if there are certain conditions in which a youth needs to observe the behavior by someone other than the person she or he admires, before acquiring it? What if specific life skills are also necessary? If such complexity is not represented in the theoretical statements describing youth behavior change, then it is implicitly assumed that those conditions are not part of the critical mass to achieve the desired change. Not only are implicit assumptions unstated, they are unquestioned, and therefore may be unrealistic and/or untested. The presence of many implicit assumptions in the theory's description of essential parts of a phenomenon is a warning sign of the theory's potential weakness.

Assumptions and a Theory's Worth

As discussed above, a theory's power is adversely affected by an insufficient number of stated assumptions. Is a theory also affected by the quality of those stated assumptions? There have been arguments on whether, since assumptions are the foci of theory, the validity of those assumptions would affect the applicability of a theory. Kies (1995) argues that relevancy and validity are twin measures of an assumption's worth. According to Friedman (1953) (in Nagel (1963)), the assumptions of a satisfactory theory could be "descriptively false" or "unrealistic," so that it is pointless to assess the merits of a theory by asking whether or not its assumptions are explicit or realistic. The relevant question is whether or not the theory yields predictions that are "sufficiently good approximations for the purpose at hand." However, it is also arguable that if a theory makes sufficient approximation of the purpose at hand, then it is making more realistic assumptions. Nagel (1963) argues against Friedman's assertion that a theory whose assumptions are in this sense unrealistic for a given domain is simply inapplicable in that domain, though it may be applicable in another. "But what is to be said of a theory whose assumptions are ostensibly unrealistic for every domain?" (Nagel 1963, p. 214). According to Nagel (1963, pp. 214–215), an assumption may be unrealistic in at least three senses

(a) An assumption can be said to be unrealistic, because it does not give an "exhaustive" description of some object, so that it mentions only some traits actually characterizing the object but ignores an endless number of other traits also present.

(b) An assumption may be said to be unrealistic, because it is believed to be either false or highly improbable on the available evidence. "The distinction between an assumption and its implications is a sharp one only in a given formulation of a theory—an implication of some assumption in one formulation may in another formulation be a premise implying that assumption" (Nagel 1963, pp. 214–215).

(c) An assumption is unrealistic when it applies to ideal conditions that are distant from actual experience or real-world situations. "Such theoretical statements (or 'laws') formulate relations specified to hold under highly 'purified' conditions between highly 'idealized' objects or processes, none of which is actually encountered in experience. For example, the law of the lever in physics is stated in terms of the behavior of absolutely rigid rods turning without friction about dimensionless points; similarly, a familiar law of pricing in economics is formulated in terms of the exchange of perfectly divisible and homogenous commodities under conditions of perfect competition. …Such statements may be said to be unrealistic but in a sense different from the two previously noted. For they are not distinguished by their failure to provide exhaustive descriptions, nor are they literally false of anything; their distinguishing mark is the fact that when they are strictly construed, they are applicable to nothing actual. According to Nagel, lack of realism can

sometimes be established on the basis of a 'directly perceived descriptive inaccuracy,' but in general, statements can be shown to be false only 'indirectly,' by first deducing from them some of their logical consequences (or implications), and then comparing the latter with 'directly' observed matters of fact" (Nagel 1963, pp. 214–215).

Friedman's line of argument is supported by Kanazawa (1988). Kanazawa argues that when a theory's assumptions are not realistic and accurate descriptions of a certain part of the empirical world, it does not mean that one cannot use the theory to explain phenomena in that part of the empirical world. From this perspective, assumptions are always empirically inaccurate as descriptive statements because they are necessarily simplifications. One does not purport to make assumptions unrealistic for its own sake; it is just an unavoidable consequence of simplification, an attempt to describe a complex phenomenon.

When certain empirical phenomena fall outside a theory's scope and conditions, one cannot use the theory to explain them, and one cannot use them to confirm or refute the theory. Kanazawa (1988) distinguishes a theory's assumptions from its scope conditions. Scope conditions are universal statements that define the class of circumstances to which a theory applies. Scope conditions define the theory's empirical range of applicability; assumptions do not (Kanazawa (1988). He argues that, nonetheless, there are conditions that would make it necessary to revise a theory's assumptions (whether unrealistic or not); when the set of assumptions leads to hypotheses that are not supported by empirical evidence. However, since hypotheses are simply postulates of assumptions and the two are thus intricately connected; it is unlikely that a set of unrealistic assumptions can generate realistic hypotheses.

Explicating Assumptions

Assumptions are the foci for any theory and thus any paradigm, as discussed above. It is also important that assumptions are made explicit in a way that realistic hypotheses can be generated and that the number of assumptions is sufficient to describe the phenomena at hand. Explication of assumptions is even more crucial in research methods used to test theories. As Bonoma (Mitroff and Bonoma 1978, p. 235) has eloquently put it: "The power of an experiment is only as strong as the clarity of the basic assumptions which underlie it. Such assumptions not only underlie laboratory experimentation but social evaluation research as well."

The notion of explicit assumptions is highly pertinent to program theory and theory-based evaluation, where it is important to deepen clarity of program design and clarify program performance. It is not uncommon for programs to have logic models that are reflections of their major theory(ies). These logic models always have assumptions elaborated, but these assumptions often pertain to factors in the external environment that would influence the success of the programs. However,

the assumptions on how program outputs are expected to translate into intended outcomes are often not explicit. It is not that these assumptions are not made. They are taken for granted, taken to be common sense, or based on "experience, intuition, and professional rules of thumb" (Weiss 2000 in Mason and Barnes 2007, p. 161). To design a project that has a good chance of success, project planners need to articulate what they expect to be true, so they and their colleagues can highlight any gaps in the logic of the program and assess whether an assumption will, in fact, turn out to be valid (Kaplan and Garrett 2005). In order to deepen understanding of program performance, a reflection on assumptions—an activity that should have happened during design, is pertinent to include early on.

According to Weiss (1995, p. 71), the explication of assumptions is a key benefit of theory-based evaluation:

> it asks program practitioners to make their assumptions explicit and to reach consensus with their colleagues about what they are trying to do and why. Without such a conversation, it is likely that different participants have different tacit theories and are directing their attention to divergent—even conflicting–means and ends. Imagine, for example, a preschool teacher who believes in unconditional affection and nurturance for the children in her care, working under a supervisor who requires that the children demonstrate cognitive achievement (numbers, colors) before they can receive approval. At the extreme, the assumptions and practices of the teacher and the supervisor may be so divergent that their efforts tend to cancel each other out.

The scale of articulation of assumptions is represented in Fig. 7.1. The following are five categories of assumptions on a continuum of articulation or explication;

(a) Very ambiguously tacit assumptions held in divergent beliefs. These may be based on experiences of individual stakeholders. For example, community members may have different interpretations of what it means to be poor, rich, lucky, or unlucky; different stakeholders may perceive well-being according to the circumstance they individually live in. These are often not expressed, and individuals may not even be aware that they hold such assumptions.

(b) Tacit but more obvious assumptions, where parties more or less have shared though unexpressed perceptions or beliefs such as those discussed in Chapter 6 under legal assumptions in litigation. Another example outside litigation could be that of some pastoral communities where burning bushes prior to a rainy season is a common and accepted practice. While everyone may believe that this practice may boost pastures, farmers may have different ideas of how this happens.

(c) Informally, explicit assumptions, e.g. indirectly expressed or inherent in shared stories, norms, uncodified symbols, and myths. Stakeholders may pursue a lengthy consultative planning process because "two (or several) heads are better than one."

(d) Assumptions that are made explicit. For instance, theories and models, policies, and programs, research and methodology all may contain specifying statements that directly or indirectly spell out assumptions. The assumptions
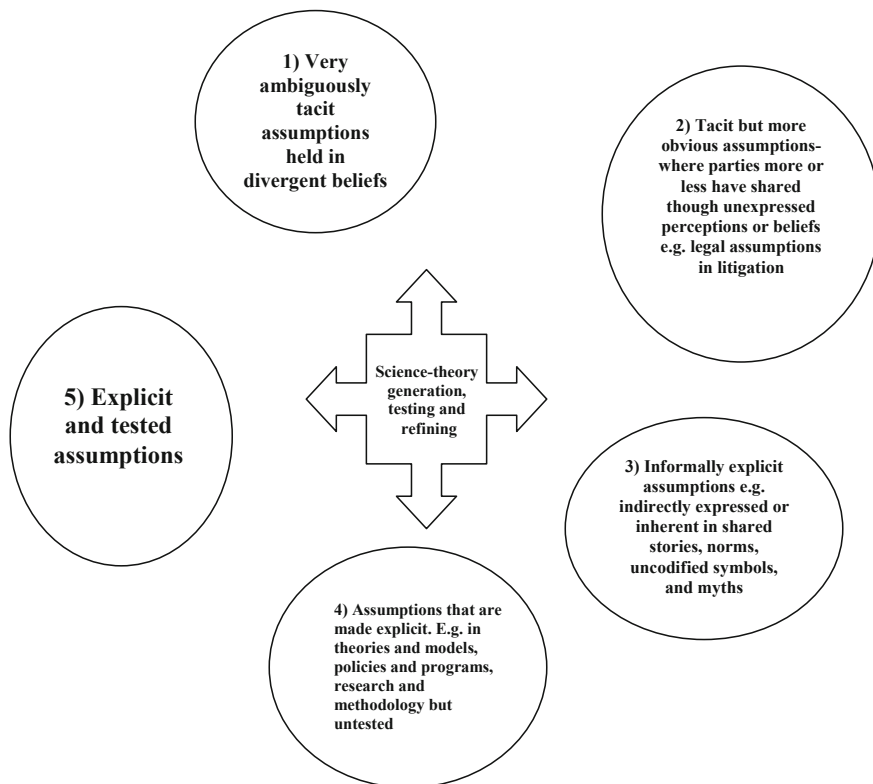
Fig. 7.1 Degrees of assumptions articulation

may be untested see the untested assumptions of community development (CD) outlined in Chap. 6 (Box 6.1). Among the assumptions—all communities know their own needs: these needs only await discovery, possibly with the assistance of a CD worker. It is further assumed that the CD worker will remain objective and lead the community to discover what it really wants (the real needs vs. the felt needs). This assumption concerning the objectivity of the CD worker rests on functional specialization and separation of roles (Kotze 1987). These characteristics are not necessarily found in developing countries where structures are functionally diffuse and where social pressures on people are experienced much differently by them than in a post industrialized country (Kotze 1987). Thus, the CD worker might find it impossible to be objective in relation to her/his job and the community she/he serves.

(e) Explicit and tested assumptions. These may have been tested by empirical means or through experience or critical reflection. A community development program in rural Albania aimed at improving well-being in the community by provision of education and health services. The assumption was that provision

of services would make the community more attractive and this would reduce migration to urban areas. However, program evaluation revealed that education resulted in more rural–urban migration as educated individuals could more easily find jobs in the urban areas (Trandafili, Personal communication). In this case, an assumption was made explicit and tested, which led to a program redesign.

The sequence or continuum of these assumptions represents different stages of theorizing or for that matter policy or program development (since a program is a theory). For example, a program's theory may be a combination of several stakeholder assumptions and beliefs that may initially be tacit, but are eventually made explicit by a skillful design facilitator. With further implementation, research and evaluation, assumptions are tested, program theory is refined, new assumptions are generated, and this healthy cycle strengthens the overall initiative. The cycle may not follow all steps on the continuum. For example, some assumptions and beliefs may be discarded in a simple discussion with other stakeholders, or by simply making them explicit—subjecting them to a logical test—hence they do not require empirical examination.

References

Babbie, E. R. (2007). *The practice of social research* (11th ed.). Belmont, CA: Thompson Wadsworth.

Becker, J., & Niehaves, B. (2007). Epistemological perspectives on IS research: A framework for analyzing and systematizing epistemological assumptions. *Information Systems Journal, 17*, 197–214.

Bennett, G. E. (1933). Assumptions. *The Accounting Review, 8*(2), 157–159.

Bhaskar, R. (1975). *A realist theory of science*. Leeds: Leeds Books.

Bonell, C. (1999). Evidence based nursing: A stereo-typed view of quantitative and experimental research could work against professional autonomy and authority. *Journal of Advanced Nursing, 30*(1), 18–23.

Borg, W., & Gall, M. (1983). *Educational research: An introduction* (4th ed.). New York: Longman Publishers.

Bredo, E., & Feinberg, W. (1982). *Knowledge and values in social and educational research*. Philadelphia: Temple University Press.

Bryman, A. (1988). *Quantity and quality in social research*. London: Routledge.

Bryman, A. (1998). Quantitative and qualitative research strategies in knowing the social world. In T. May & M. Williams (Eds.), *Knowing the social world*. Buckingham and Philadelphia: Open University Press.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*, 97–113.

Campbell, R., & Wasco, S. M. (2000). Feminist approaches to social science: Epistemological and methodological tenets. *American Journal of Community Psychology, 28*(6), 773–791.

Carter, S. M., & Little, M. (2007). Justifying knowledge, justifying method, taking action: epistemologies, methodologies, and methods in qualitative research. *Qualitative Health Research, 17*(10), 1316–1328.

Clark, A. M. (1998). The qualitative-quantitative debate: moving from positivism and confrontation to post positivism and reconciliation. *Journal of Advanced Nursing, 27*, 1242–1249.

Cook, J. A., & Fonow, M. M. (1986). Knowledge and women's interests: Issues of epistemology and methodology in feminist sociological research. *Sociological Inquiry, 56*, 2–29.

Coomer, D. (1984). Critical science: Approach to vocational education research. *Journal of Vocational Education Research, 9*(4), 34.

Creswell, J. W., Hanson, W. E., Plano, V. L., & Morales, A. (2007). Qualitative research designs: Selection and implementation. *The Counseling Psychologist, 35*(2), 236–264.

Gioia, D. A., & Pitre, E. (1990). Multiparadigm perspectives on theory building. *The Academy of Management Review, 15*(4), 584–602.

Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: from intuition to institution. *The Academy of Management Review, 24*(3), 522–537.

Crossan, M., & Guatto, T. (1996). Organizational learning research profile. *Journal of Organizational Change Management, 9*(1), 107–112.

Crossan, M., Lane, H., White, R. E., & Djurfeldt, L. (1995). Organizational learning: Dimensions for a theory. *International Journal of Organizational Analysis, 3*, 337–360.

Garrick, J. (2010). Doubting the philosophical assumptions of interpretive research. *International Journal of Qualitative Studies in Education, 12*(2), 147–156.

Granello, D. H. (2002). Assessing the cognitive development of counseling students: changes in epistemological assumptions. *Counselor Education and Supervision, 41*(4), 279–293.

Greene, J. C., & Curucelli, V. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation, 74*(Summer), 5–17.

Greene, J. C., Lipsey, M. W., & Schwandt, T. A. (2007). Method choice: Five discussant commentaries. *New Directions for Evaluation, 113*(Spring), 111–118.

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

Hammersley, M. (1995). *The politics of social research*. London: Sage.

Hawkins, D. F. (1978). Applied research and social theory. *Evaluation Quarterly, 2*(1), 141–152.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. *Qualitative Market Research: An International Journal, 3*(3), 118–126.

Hjørland, B. (2002). Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Technology, 53*(4), 257–270.

Hofer, B. K., & Pintrich, P. R. (2002). *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah: Erlbaum.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation, 52*(1), 3–50.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast, ACM transactions on computer. *Human Interaction, 3*(4), 320–351.

Kanazawa, S. (1988). In defense of unrealistic assumptions. *Sociological Theory, 16*(2), 193–204.

Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning, 28*, 167–172.

Kerem, E., Fishman, N., & Josselson, R. (2001). The experience of empathy in everyday relationships: Cognitive and affective elements. *Journal of Social and Personal Relationships, 18*(5), 709–729.

Kies, D. (1995). Assumptions. Retrieved July 12, 2011 from http://papyr.com/hypertextbooks/comp2/assume.htm.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Lynn, S. J., Weekes, J. R., & Milano, M. J. (1989). Reality versus suggestion: Pseudo memory in hypnotizable and simulating subjects. *Journal of Abnormal Psychology, 198*(2), 137–144.

Mason, P., & Barnes, M. (2007). Constructing theories of change: Methods and sources. *Evaluation, 13*(2), 151–170.

Mingers, J. A. (2003). Classification of the philosophical assumptions of management science. *Journal of the Operational Research Society, 54*(6), 559–570.

Mir, R., & Watson, A. (2001). Critical realism and constructivism in strategy research: Toward a synthesis. *Strategic Management Journal, 22*, 1169–1173.

Mitroff, J., & Bonoma, T. V. (1978). Experimentation, and real world problems. A critique and an alternate approach to evaluation of psychological assumptions. *Evaluation Quarterly, 2*(2), 235–260.

Morgan, G., & Smircich, L. (1980). The case for qualitative research. *The Academy of Management Review, 5*(4), 491–500.

Nagel, E. (1963). Assumptions in economic theory. *The American Economic Review, 53*(2), 211–219.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: Guilford Press.

Perry, C. (1998). Processes of a case study methodology for postgraduate research in marketing. *European Journal of Marketing, 32*(9), 785–802.

Robson, C. (1993). *Real world research: A resource for social scientists and practitioners-researchers*. Oxford: Basil Blackwell.

Sale, J. E., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Revisiting the Quality and Quantity, 36*, 43–53.

Scates, D. E. (1940). Assumptions underlying research data. *The Journal of Educational Research, 34*(4), 241–254.

Scott, D., & Usher, R. S. (1996). *Understanding educational research*. London: Routledge.

Shermer, M. (2011). *The believing brain: From ghosts and gods to politics and conspiracies-How we construct beliefs and reinforce them as truths*. New York: Times Books.

Shrivastava, P., & Mitroff, I. I. (1984). Enhancing organizational research utilization: The role of decision makers' assumptions. *The Academy of Management Review, 9*(1), 18–26.

Schwandt, T. A. (2001). *Dictionary of qualitative inquiry* (2nd ed.). Thousand Oaks: Sage.

Shah, S. K., & Corley, K. G. (2006). Building better theory by bridging the quantitative–qualitative divide. *Journal of Management Studies, 43*(8), 1821–1835.

Tsoukas, H. (1991). The missing link: A transformational view of metaphors in organizational science. *Academy of Management Review, 16*(1), 566–585.

Ullman-Margalit, E. (1983). On presumption. *The Journal of Philosophy, 80*(3), 143–163.

Warner, R. S. (1978). Toward a redefinition of action theory: paying the cognitive element its due. *American Journal of Sociology, 83*(6), 1317–1349.

Weiss, C. H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. In K. Fulbright-Anderson, A. C. Kubrisch, & J. P. Connell (Eds.), *New approaches to evaluating community initiatives: Theory, measurement and analysis* (Vol. 2, pp. 65–92). Washington, DC: Aspen Institute.

Weiss, C. (2000). Which links in which theories shall we evaluate? *New Directions for Evaluation, 87*(Fall), 35–45.

Wegerif, R. (2008). Objective approaches: Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal, 34*(3), 347–361.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.

Williams, D. D. (1986). When is naturalistic evaluation appropriate? *New Directions for Program Evaluation, 30*(Summer), 85–92.

Yauch, C. A., & Steudel, H. J. (2003). Complementary use of qualitative and quantitative cultural assessment methods. *Organizational Research Methods, 6*(4), 465–481.

Part III
Explicating Tacit Program Assumptions

Chapter 8
Diagnostic Assumptions

Abstract Diagnostic assumptions are stated as stakeholders' expectations or beliefs of the major and minor causes of core problems. Since the intervention to address a problem is based on the causes of that problem, diagnostic assumptions are crucial to a normative theory and need to be examined from design, implementation, and evaluation perspectives. This chapter appraises the use of the *policy scientific approach* in explicating diagnostic assumptions. An Alternative Causes Approach is proposed.

> *We believe it is fundamentally important for us to examine the assumptions underlying our ways of doing things and our attitudes toward educational objectives and processes. Some of these attitudes were acquired by mere gregarious assent. Now that we have better means of checking up on our assumptions, we should proceed to overhaul the whole question of organizing, managing, and teaching our public schools. This is a big job—too big to be undertaken all at once. Under these circumstances the most important things should be done first. We suggest that they are not the minutiae but the broad basic facts underlying the larger aspects of education. In short, the time has come when we may properly examine the assumptions which we have considered sufficient and see whether, with the existing means of measurement and analysis, these assumptions maintain their validity*
>
> Educational Research Bulletin, editorial; 1923, p. 276

Diagnostic assumptions
Problem assessment:
identifying the root/major
causes of the problem on
which to base an intervention
strategy, which other causes
may be taken for granted?

Prescriptive assumptions:
Once a core problem cause has
been identified; in selecting a
strategy or intervention to
address the problem and
causes; what other alternatives
may or consideration may be
taken for granted?

Transformational assumptions:
How are the intervention's
immediate outputs expected to
translate into desired outcomes
and goals (results)? What
conditions may be taken for
granted

Fig. 8.1 Assumptions in program design

Explicating Tacit Program Assumptions

Chapters 9 and 10 (also this chapter) discuss three categories of tacit program assumptions and suggest ways of making these assumptions explicit. These three tacit assumption categories include:

(a) Diagnostic assumptions
(b) Prescriptive assumptions
(c) Transformational assumptions

These are expectations, beliefs, or propositions commonly taken for granted, and thus they are seldom stated explicitly in program documents. They revolve around three of the major phases of program cycle: (a) problem analysis/needs assessments (diagnostic assumptions, (b) choice or selection of interventions or program strategies (prognostic assumptions), and (c) setting program goals or

outcomes (transformational assumptions). These three categories are highlighted in a modified program cycle Fig. 8.1.

Note that this is a modified program cycle. This cycle does not address planning for and conducting program implementation, monitoring or evaluation. All these phases involve many assumptions similar to the three categories discussed here. The assumptions made in other phases (implementation, monitoring, and evaluation) are outside the scope of this discussion. The factors discussed here are those too often taken for granted at the early stages of building a program's theory. Yet they are foundational to success. If a program's theory is underpinned by "faulty" assumptions, implementation and program outputs are not likely to yield expected results and monitoring and evaluation information may be misleading.

These three chapters demonstrate that several approaches in theory-based evaluation attempt to bring these assumptions to light, but often during evaluation rather than at the point when programs are being formulated or designed. Under these circumstances, evaluators—who tend to ask these difficult questions—do not engage with the programs until it is time to evaluate them. This discussion assumes that there is an evaluative reflection within the program's design that unearths these assumptions before the evaluation phase, when it cannot address program design and execution.

Equally important, these foundational assumptions are at the heart of the program. If a problem is designed on the basis of faulty assumptions, the strategy(ies) are likely to be flawed, and the intervention will not be on target—indeed the results may fall far outside the intended outcome(s). For example, in a community where most of the children of school age are out of school, stakeholders may feel that the major cause of low enrollment is the lack of schools in the locality. In taking this position, a number of considerations could be implied:

(a) The available schools cannot enroll any more children and this may have affected the quality of education.
(b) Children have lost interest in going to school or parents/caregivers are demotivated from taking children to school.
(c) Where schools are not congested or overcrowded, the quality is good and probably children and caregivers are interested.
(d) Children closer to schools are more likely to be enrolled than those further away from the school.
(e) Caregivers/community members value their children's education and are simply limited by the small number of schools.

Often, reflection on problem causes does not go deep enough to question such "implied possibilities", which if explicitly stated, could be subjected to a test—perhaps a simple logical analysis that does not require experiment or by the application of evidence and experience of key community stakeholders.

What if there are few schools, but caregivers are just not keen enough to take children to school? Or, what if most caregivers think the roads to schools are not entirely safe? Or what if the children closest to the school are not the most likely to be enrolled? Any of these conditions might result in a program built on faulty

assumptions and generating inappropriate solutions. Inevitably, it would turn out that more schools would not necessarily elicit greater enrollment, or greater literacy—not unless the actual causes and problems could be isolated.

Diagnostic Assumptions

Diagnostic assumptions are stated as stakeholders' perceptions of the major and minor causes of the core problems. Since the intervention to address a problem is based on the causes of that problem, diagnostic assumptions are crucial to a normative theory and need to be examined from design, implementation, and evaluation perspectives. As defined in Chap. 5, normative theory refers to what the structure of the program should be—the intervention, implementation processes, and outcomes (Chen 2005).

Diagnostic assumptions are most commonly made during the early phases of program planning—particularly the needs assessment/problem identification stage. The Merriam-Webster dictionary defines diagnosis as the "investigation or analysis of the cause or nature of a condition, situation, or problem". Within a typical program cycle, diagnostic assumptions arise during problem analysis or needs assessment. Needs assessment is a diagnostic activity; it also involves identifying resources that must be addressed and opportunities or potential that can be harnessed to address the problems identified (Wandersmana et al. 2000). Good needs assessment is at the heart of designing effective programs. The more thoroughly the specific needs of a target population within a specific context are understood, the better the eventual outcome and the more effective a program will be (Berberet 2006). Involvement of target population and program stakeholders logically increases the efficacy of program design (Berberet 2006). It involves three steps:

(a) Identification of the stakeholders affected by the proposed project.
(b) Identification of the major problems faced by beneficiaries and the opportunities that exist to address the problems.
(c) Development of a problem tree to establish causes and effects (EC 1999a, b).

It is worth noting that a key assumption often made in participatory needs assessment processes (where community members vote on problems and causes) is that the needs or causes with the greatest number of votes represents the most pressing needs or most serious causes of the problem. The risk to this assumption is that "felt" needs may not necessarily be the real needs of the community. Also, stakeholders understanding of needs/problems and their causes is likely to be based on different beliefs and ideas. These beliefs or experiences are seldom explicitly stated during the program planning process. They are sometimes "discovered" during program evaluation. There are a number of evaluation reports that have unearthed initially tacit diagnostic assumptions. Below are some examples:

Lozare (2010) evaluated an HIV prevention program in which it was assumed that people are infected with HIV because they are not aware of the risks and how

to avoid them, or even when they are aware, they do not have the life skills (skills for personal management and social interaction necessary for adequate functioning on an independent basis, see Ouane 2002) to overcome them. The solution would be awareness and increasing knowledge to change behavior and attitudes. Thus, the more individuals understand risky situations, the more they avoid them and exposure to HIV, which assumes that they have resources needed to behave in the ways desired.

Marcano and Ruprah (2008) studied Chile's Progressive Housing Program, PHP; a public housing program that finances new houses. This program aimed at providing housing to the poor. The eligibility requirements for targeting the poor—included a certain level of savings by the home purchaser, implying that some assumptions had been made about why the poor did not have enough housing, and/or about the characteristics of poor people. One such assumption could have been that the poor had some income that was available to be saved or set aside. Another could have been the assumption that a savings culture involving financial institutions even existed. The findings of the evaluation show that the program was able to find people who have a savings culture, as well as meeting other criteria. But these were not the poor that the program was intended to reach. Could this failure have been because of faulty diagnostic assumptions? Quite possibly.

Kruse-Levy et al. (2007), evaluated a program that was assisting people living with HIV and AIDS who had been marginalized due to illness to reintegrate socially and economically into society after regaining their health through anti-retroviral therapy. The program provides group and family counseling, basic training for managing a small business, apprenticeships, job placements, vocational training, small grants, health education, reconciliation with estranged families, and other services required for clients to transition from being dependent upon project assistance to supporting themselves. The approach taken here assumes that reintegration and change in social status are as much a function of what individuals do with their lives and/or how these individuals shape the way others would look at them after therapy. It is interesting to note that other programs supporting reintegration have assumed that social inclusion and stigma are functions of community perceptions and attitudes and have thus designed community education targeting these attitudes. The evaluation of this program recommended a balance between the two approaches.

In Makayi et al. (2006), Oxfam GB provided an unconditional, monthly cash transfer of ZK 90,000 per month (approximately 26 US dollars) up to 13,500 households (about 30 % of the rural population) in Kaoma and Mongu Districts, Western Province of Zambia during the months from November 2005 to March 2006. Two hundred households were provided with cash in Mongu District in October as a test/trial. The purpose of the project was to support households to achieve short-term food security without having to employ harmful coping strategies and to prevent malnutrition. Community committees in each ward were responsible for identifying the poorest households and lists came out of each ward, compiled as project beneficiaries. Inter-ward variation in need may not have been

taken into account, and this assumption may create wide disparities among the poor at district level. Evaluation showed that from a nutritional perspective, it seems likely that the majority of beneficiaries would have been able to meet minimum food needs and there would not have been significant increase in acute malnutrition even without the cash assistance. This was due to inexplicit assumptions in assessing—e.g., that the precondition for improved nutrition was cash transfers, or that the beneficiaries selected were not the initially conceived poor. Consumption patterns, availability and diversity of food, utilization of cash, and intra- household relationships were assumed as crucial preconditions although they were never monitored. Also, it was assumed that the major proportion of household expenditure is on food for household consumption, specifically on maize meal, but also on cassava, oil, and fish and that some households are spending a small proportion of money on food to share with non-household members. This assumption was verified at assessment and mid-term evaluation. Also assumed was that women should receive and control the money given to the household, since they are responsible for buying food. This also assumes that the gender division of labor is supportive and additional education and sensitization for men was provided to help the wives more readily control the money for food.

All these evaluations (and several others), highlight several assumptions initially made about the actual causes of the problem. These assumptions were not made explicit until the programs were evaluated. Is it too late to make assumptions explicit at the evaluation stage? No, it is never too late to unearth assumptions. But earlier attention might have increased program success. As suggested in Mayne's (2011) contribution analysis, unearthing "the behind the scene" assumptions during evaluation helps to clarify the programs performance story. However, another critical thing about a program's performance, other than clarifying it, is achieving or realizing it to at least the extent intended. And if a program's theory is built on the wrong assumptions about the causes of the problem a program is attempting to address, it decreases the opportunity to generate the desired performance, and it complicates linking program outputs with program inputs in the evaluation phase.

Explicating Diagnostic Assumptions

The Policy Scientific Approach

The policy scientific approach (PSA) is an approach that has been used to spell out diagnostic assumptions. Leeuw (2003: pp. 8–9) in his article "Reconstructing program theory: a policy-scientific approach" examines the use of the PSA to refine or develop a theory for a program that is already in place. This approach, as explained by Leeuw, can be useful in helping stakeholders make clear what otherwise might be tacit assumptions during a problem diagnostic process.

The approach has six steps.

- Use secondary data (gathered through program-related documents) and primary data (such as interviews with stakeholders) to identify behavioral mechanisms expected to solve the problem. This activity should address why it is believed necessary to solve the social, organizational, or policy problem; and what the goals are of the policy or program under review. Leeuw argues that the nature of these statements is particularly important:

  It is evident that x… will work,

  In our opinion, the best way to go about this problem is to…,

  The only way to solve this problem is to…,

  Our institution's x years of experience tells us that….

  Leeuw (2003, pp. 8–9)

- After problem-related statements are identified, a survey is compiled to link these statements with the intervention mechanisms and the program or policy goals.
- The problem-related statements are reformulated in conditional "if–then" propositions or propositions of a similar structure ("the more x, the less y").
- The missing links between statements or propositions are addressed with warrants (the "because" part of an argument) e.g. "The organization's performance will not improve next year" follows from "The performance of this organization has not improved over the last 5 years" because of the principle, "past performance is the best predictor of future performance." Leeuw (2003, pp. 8–9). The "because" part of such an argument is often not made explicit. Consequently, these warrants must be inferred by the person performing the analysis.
- The warrants are reformulated in terms of conditional "if–then" (or similar) propositions and a chart is drawn showing the (mostly causal) links.
- The validity of the propositions are evaluated for (a) logical consistency of the set of propositions; (b) their empirical content, that is, the extent to which the theory and in particular the assumed impact of the behavioral mechanisms correspond with the state-of-the art within the social/behavioral/economic sciences on these mechanisms; and (c) the extent to which the program's theory focuses on variables that it can have an effect on. And that can be "manipulated" or "steered" through policy programs.

The PSA is becoming a popular approach in evaluations, especially where there is a lack of clear description of a program's theory. van Noije and Wittebrood (2010) have used the PSA to examine the effectiveness of policy interventions to fight crime and the validity of the policy theory that underlies them. In using this approach, the lack of clearly articulated policy assumptions becomes even more apparent: "It quickly became clear that these notes hardly make explicit which

assumptions underpin the policy choices." (van Noije and Wittebrood 2010, p. 499).

Kautto and Similä (2005) use the PSA to evaluate the Finnish Environmental Protection Act, 1999. The reform integrated five different permits (air pollution, water pollution, waste management, protection of health, and neighborhood relations) into one environmental permit. The intervention theory of integrated permits was constructed using both European and national legislative documents. In these documents the benefits of integrated permits were defined in relative terms: integrated permits were assumed to be superior in comparison with a system of several sectoral permits.

On the basis of the documents it was possible to identify assumed chains of influence (i.e. assumptions) on which the superiority was based. The assumptions were the following.

1. Integration of permits will abolish the gaps between different permits.
2. Detrimental side-effects of environmental measures will be reduced, because pollution will no longer be transferred from one environmental medium (water, air, soil) to another.
3. Integration will enable a new kind of prioritization.
4. Integration of permits will result in technological change, from 'end-of-pipe' technologies to process technologies.

After the formulation of the intervention theory of IPPC permits as part of the new regulations, it became possible to ask what kind of information was available to scrutinize whether theory-based assumptions hold true.

With regard to the first assumption, the outputs of administration (i.e. permits) provided a good basis to make observations. In fact, the essence of this assumption was that the IPPC permits would differ from the old permits (the gaps would be filled because the competencies of public authorities were extended). Thus, the evaluators could formulate a more precise question: do permits contain new kinds of provisions that regulate environmental problems that were not regulated before? On the basis of interviews with stakeholders, it was concluded that the conditions set had changed very little. Thus, there was only weak evidence that this assumption of efficacy held.

With respect to the second assumption, evaluators formulated a measurable question: have the authorities taken account of side-effects like transferring emissions from one environmental medium to another as an argument when designing the content of the permits? On the basis of permit analysis as well as interviews, the answer was negative; that is, the second assumption did not hold. In addition, it was possible to approach the issue of setting priorities on the basis of the decisions, because permit authorities are obliged to justify their decisions, and this aspect certainly was one that needed to be justified. The permit documents confirmed that the argument had not been used.

Additionally, the issue was cross-checked in the interviews, and this confirmed the result, with a modification: most of the interviewees were of the opinion that priority setting had not affected the design of permits. To make the intervention

theory more complete, two additional assumptions—or reformulated warrants, using Leeuw's terms—had to be formulated on the basis of an analytical reconstruction of the intervention theory:

(a). Technological change will occur only if the activities concerned have impacts on more than one environmental media (air, water, land); and

(b). Technological change is linked to the investment cycles of the plant concerned and regulatory authorities have only minor possibilities to affect these cycles.

Ehren et al. (2005) used the PSA to analyze assumptions that underpinned the inspection function of the Dutch Primary Education system. In unearthing these assumptions, Ehren et al. reconstructed a program theory for education inspection. The assumptions explained how education inspection would lead to school improvement. For example, one of the statements/propositions framed about education was that the existing level of education quality in schools had a lot to do with compliance control. Such a statement translated into the following assumption: "If the inspectorate controls compliance of schools, then schools will (eventually) attain a satisfactory level of quality."

Another proposition was that over-inspection among other factors might have undermined schools' sense of responsibility for enhancing school quality. Schools might have believed that quality is the duty of the inspectorate. And when they are rewarded with less intense and less frequent inspection, schools will be motivated to develop good quality assurance systems. These propositions translated into the following assumptions:

- If the inspectorate decreases the intensity and frequency of inspection in a case where the school has developed a quality assurance system and delivers good educational quality, schools will develop systems for quality assurance.
- If the inspectorate employs proportional inspection, schools will offer more added values.
- If the inspectorate publishes its findings in an accessible manner, schools will offer more added values.

In unearthing these assumptions, Ehren et al. argue that evaluation of these assumptions was the basis for predicting the effectiveness of the inspection legislation. In addition, unearthing these assumptions can be useful in improving the design of the program in advance of implementation and evaluation, which can in turn lead to substantial savings in time and resources (Fitzpatrick 2002). Donaldson elaborates this very articulately:

> It is very common to discover the program under investigation is poorly conceptualized. For example, when you take a presumed outcome and examine it carefully, it is common to find that the program as designed has very little chance, based on what we know from prior research and evaluation, to account for a significant portion of the variance in the presumed outcome. It is also common to discover that the relationship between a program and its presumed outcomes may be conditional or better explained by taking moderating factors into account (In Fitzpatrick 2002, p. 347).

However, while unearthing assumptions helps to improve a program's theory; a correct, realistic, and implemented program theory does not guarantee success, as other factors may play a role (Leeuw 2003). Other than unearthing assumptions "retrospectively" in reconstructing program theory, there is even more value in unearthing tacit assumptions at the very early stages of formulating a program theory and in identifying problem causes.

The Alternative Causes Approach

A number of tools such as 'problem tree analysis' are used to facilitate stakeholders' work in problem assessment. Popularized by LFA, the problem tree analysis has been one of the most commonly used needs assessment and situational analysis tools (e.g., EC 1999a, b, 2002).

According to Social Impact (1999), problem tree analysis (also called situational analysis or just problem analysis) helps to find solutions by mapping out the cause and effect anatomy around an issue in a similar way to a Mind Map, but with more structure. The approach has several advantages (Social Impact 1999):

- The problem can be broken down into manageable and definable chunks. This enables a clearer prioritization of factors and helps focus objectives.
- There is more understanding of the problem and it is often interconnected and even contradictory causes. This is often the first step in finding win-win solutions.
- It identifies the constituent issues and arguments, and can help establish who and what the political actors and processes are at each stage.
- It can help establish whether further information, evidence or resources are needed to make a strong case, or build a convincing solution.
- Present issues—rather than apparent, future, or past issues—are dealt with and identified.

The process of analysis often helps build a shared sense of understanding, purpose, and action.

In the problem tree analysis, the problems set out in a hierarchical order. First, each identified problem is summarized. From these a starter problem is selected, and a second problem related to it, then: "if the problem is a cause it goes on the level below"; if it is an effect it goes above; if it is neither a cause nor an effect it goes on the same level.

For example, if the focal problem is "poor nutritional status of babies and infants", a cause might be "poor availability of high protein foods", while an effect might be "high rates of infection among babies and infants".

As the tree develops, the remaining problems are attached to it in the same way. Once the problem tree is complete, a focal problem is selected. The focal problem should be agreed on by the different interest groups as being the central problem
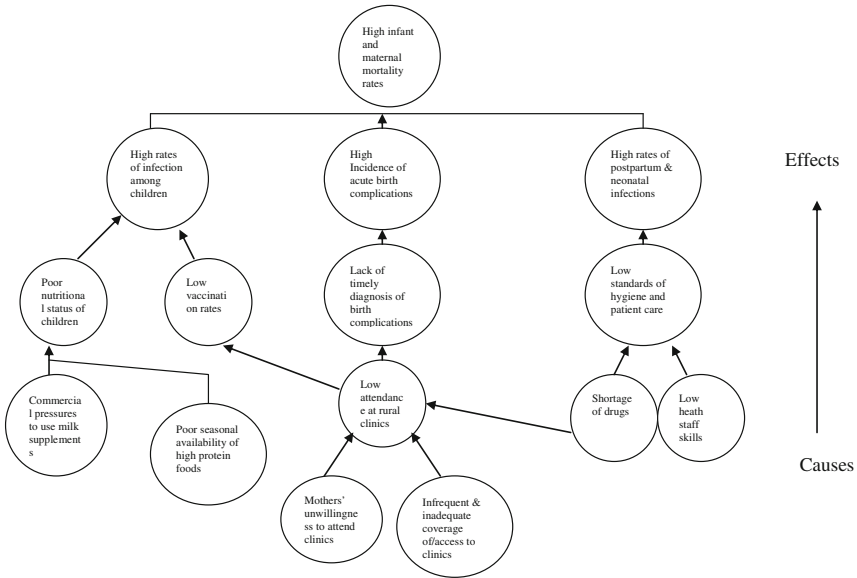
Fig. 8.2 A problem tree for a maternal and child health program (adapted from EC 1999a,b, p. 18)

that should be addressed by the project or intervention (EC 1999a, b) (See Fig. 8.2). This is usually a very messy and iterative process.

It is essential that, to the extent possible, problem causes are accurately assessed. There are always several other alternative causes to a problem. Sometimes these are identified within the discussion, and sometimes they may not be identified. It is important that the facilitator help stakeholders to reflect analytically and identify the most plausible causes of the problem.

There are problem causes identified as plausible by stakeholders and those that are potentially possible but perhaps not in this environment or circumstance and those that are just not thought of. The facilitator needs to guide stakeholders on these dimensions of causes.

(a) **What are the major causes (very certain or less certain?)**
For example, in Fig. 8.2 above, the two causes of poor nutrition identified above are the lack of protein rich foods and commercial pressures to use milk supplements. These could be related to other causal factors as well, such as lack of knowledge of appropriate feeding practices, or inadequate or unclean water supply.

(b) **What are the minor causes (very certain or less certain?)**
Stakeholders may consider other 'minor' factors to which the core problem may be attributed e.g. may be care givers cannot afford the right food, or caregivers do not have the food.

Table 8.1 Alternative causes matrix

	Stakeholders' (dominant) beliefs on the major cause (s) of the problem	What experience from similar environments or the same community shows to be the major cause (s) of the problem	What evidence (perhaps empirical from applicable research) shows to be the major causes of the problem assessed? This could be from analysis of data generated from assessment or other scientific theory
Obviously Plausible	A (the obviously plausible causes are based mainly on stakeholder beliefs rather than relevant experience or evidence)	B (Causes are 'obvious' because of stakeholder experiences)	C (Causes are 'obvious' because they are based on empirical evidence)
Minimally plausible	D (the minimally plausible causes are based mainly on stakeholder beliefs rather than relevant experience or evidence)	E (the minimally plausible causes based on stakeholder experiences)	F (Causes are minimally plausible because they are based on empirical evidence)
Unthinkable (factors are thought to have no connection with the problem in this particular situation)	G (the unthinkable causes, based mainly on stakeholder beliefs rather than relevant experience or evidence)	H (the unthinkable causes based on stakeholder experiences)	I (Causes are unthinkable because they are based on empirical evidence)

(c) **What are the improbable causes (certain or not)**
  The facilitator asks stakeholders to identify those factors they think are not or could not be attributed to causing this problem. These are usually checked against current evidence or experiences or beliefs. For example, in this case, malnutrition is not caused by lack of clean water.

It is important to outline all these perspectives and expectations around problem causes. Most often problem analysis identifies what stakeholders think are the obvious causes, but this may not give a deep enough basis for planning. Elaborating the obviously plausible, minimally plausible, and the unthinkable also widens (and deepens) reflection on problem causes and is likely to get stakeholders to rethink the causes they might initially express.

The other step in analysis of problem causes is to determine the nature of information they are based on (see Table 8.1). These may include but are not limited to:

- Stakeholders' (dominant) beliefs on the major cause(s) of the problem
- What experience from similar environments or the same community shows to be the major cause(s) of the problem
- What evidence (perhaps empirical from applicable research) shows to be the major causes of the problem assessed (This could be from analysis of data generated from assessment).

The matrix above can help stakeholders review their notions of the core problem-'obviously plausible' 'minimally plausible' and 'unthinkable' based on whether it is simply the stakeholders' beliefs, experiences in similar or the same environment and the applicable evidence.

In scenario A, the obviously plausible causes are based mainly on stakeholder beliefs rather than relevant experience or evidence. This could be due to long-held cultural beliefs, which is not uncommon. Sometimes causes are 'obvious' because of experiences (B) and researched or empirical evidence (C). Whatever the basis of cause expectation may be, it is important to probe into the why. Definitely it is necessary for wrong beliefs to be tactfully challenged by evidence. It is also true that in most cases, evidence is in short supply and its quality may vary. Sometimes what is called evidence may be a few positive or negative stories or cases that do not really represent the cause. Triangulating/collaborating the three (or more) forms of problem cause expectations helps to generate a more reliable problem cause–whether it is the obvious, minimally plausible (D, E, F), or unthinkable (G, H, I).

The three categories of problem causes (and the extent of certainty as well as arguments/warrants) need to be documented. Although the intervention is most likely going to be based on the 'obviously plausible causes' of the problem, the minimally plausible and the unthinkable categories should be documented within the diagnostic assumptions. The diagnostic assumptions will state the obviously plausible, the minimally plausible, and the unthinkable causes of the problem: e.g., "The problem of child malnutrition is (obviously) caused by the lack of protein-rich foods and commercial pressures to use milk supplements, more than afford-ability/income and it has nothing to do with poor hygiene".

A water and sanitation program could determine that the high incidence of waterborne illnesses in a community is mostly because of the poor water infra-structure; to a minor extent due to poor hygiene practices, but it has nothing to do with affordability of water or income levels.

Stakeholders may elaborately state that an education program has been designed to address the problem of low enrollment which is majorly caused by lack of school infrastructure, and to a minor extent negative cultural beliefs and may have nothing to do with school management or teacher capacity.

Diagnostic assumptions can be stated in different ways. They do not, for example, have to be in one statement as the examples above; they do not have to highlight just one or two plausible causes. What is most important is that stakeholders describe a range of possible and not plausible causes.

References

Berberet, H. M. (2006). Putting the pieces together for queer youth: A model of integrated assessment. *Child Welfare, 85*(2), 361–377.

Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Newbury Park: Sage publications.

Editorial Comment. (1923). Fundamental assumptions. *Educational Research Bulletin, 2*(16), 276.

Ehren, M. C. M., Leeuw, F. L., & Scheerens, J. (2005). On the impact of the Dutch educational supervision act analyzing assumptions concerning the inspection of primary education. *American Journal of Evaluation, 26*(1), 60–76.

European Communities (EC) (1999a). *Project cycle management training handbook*. Brussels: EC.

European Commission. (1999b). *Joint relex service for the management of community aid to non-member countries (SCR). Resources, relations with the other institutions, evaluation, and information evaluation*. Sussex: EC.

Fitzpatrick, J. (2002). Dialogue with Stewart Donaldson. *American Journal of Evaluation, 23*(3), 347–365.

Kautto, P., & Similä, J. (2005). Recently introduced policy instruments and intervention theories. *Evaluation, 11*(1), 55–68.

Kruse-Levy, N., Senefeld, S., Sitha, A., & Arturo, A. (2007). *Bridges of hope socioeconomic reintegration project report of a follow-up survey with clients living with HIV and AIDS*.Catholic Relief Services (CRS): Baltimore.

Lozare, B. V. (2010). Summative evaluation surveys: Ethiopia, Rwanda and Uganda. Catholic Relief Services-Avoiding Risk, Affirming Life Program. Catholic Relief Services (CRS): Baltimore.

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation, 24*(1), 5–20.

Makayi, M., Judge, D., Nyungula, M., & Leather, C. (2006). *Evaluation of cash transfer pilot project in Western Province, Zambia*. Oxford: OxfamGB.

Marcano, L., & Ruprah, I. J. (2008). *An impact evaluation of Chile's progressive housing program*. Washington, DC: Inter-American Development Bank (Working Paper: OVE/WP-06/08).

Mayne, J. (2011). Contribution analysis addressing cause effect. In K. Forss, M. Marra & R. Schwartz (Eds.), *Evaluating the complex: Attribute contribution and beyond,* New Brunswick (pp. 53–96) New Jersy: Transactional Publishers

Ouane, A. (2002) Key competencies in lifelong learning. *Institutionalizing lifelong learning: Creating conducive environments for adult learning in the Asian context*. Hamburg: UNESCO Institute for Education.

van Noije, Lonneke, & Wittebrood, K. (2010). Fighting crime by fighting misconceptions and blind spots in policy theories: An evidence-based evaluation of interventions and assumed causal mechanisms. *American Journal of Evaluation, 31*(4), 499–516.

Social Impact (1999). *Problem tree analysis*. Retrieved November 22, 2011, from http://webarchive.nationalarchives.gov.uk/+/http://www.dfid.gov.uk/FOI/tools/chapter_03.htm

Wandersmana, A., Imma, P., Chinmanb, M., & Kaftarian, S. (2000). Getting to outcomes: A results-based approach to accountability. *Evaluation and Program Planning, 23*(2000), 389–395.

Chapter 9
Prescriptive Assumptions

Abstract Prescriptive assumptions relate to the intervention or strategy devised for the problem or to reach a stated objective, which represents stakeholders' beliefs of what could be the best ways to address the problem or need. The use of goal evaluation and strategy assessment for explicating prescriptive assumptions is appraised. The alternative strategies approach is proposed.

> *Dialogue about contested meanings cannot be about the object*
> *meanings being contested but must delve into the underlying*
> *beliefs, assumptions, and experiences which contextualize and*
> *thus signify the focal objects in multiple ways*
>
> Myers 1995, p. 582

Introduction

The term "prescriptive" emphasizes a connotation that once a core problem is identified by stakeholders a "prospective of recovery" is the intervention devised to address that problem. Prescriptive assumptions reflect what we think ought to be happening in a particular situation (Brookfield 1995; Brandenburg 2009). Thus, prescriptive assumptions relate to the intervention or strategy devised to resolve the problem or to reach a stated objective, which represents stakeholders' beliefs of what could be the best ways to address the problem or need. The process during which prescriptive assumptions are made is commonly referred to as program design. Program design is the process of planning appropriate program and project strategies using assessment results, to show how issues identified can be addressed (WVI 2007). During program design, community needs, rights, and priorities are all taken into account in deciding whether to implement the program. "The

purpose of design is to end up with a logical plan integrating prioritized issues with opportunities identified in assessment, so the resulting program and projects can be implemented and managed with accountability and for learning" (WVI 2007, p. 40). A design process results in:

- A better understanding of how to address issues identified and solid analysis of community needs, rights and responsibilities, and opportunities;
- A description of how best to leverage opportunities, and an outline of respective roles and responsibilities of communities and local government to address problems faced, as well as the organization's roles in facilitating these changes;
- An outline of program or project rationale in how it will address issues, communicating this using a logical framework; a comprehensive monitoring and evaluation plan describing how activities will be tracked and progress toward success measured; and
- A strategy for learning and accountability (WVI 2007).

The nature of problem and causes assessed, as well as assumptions made about the problem(s) and cause(s) inevitably influence or determine how stakeholders think of addressing this problem. According to House (1979) the framing of the social problems is critical to the solutions that emerge. For example, if it is assessed that a problem of poverty is caused by fragmented social services, the implicit solution is that social services should be better coordinated. Korten (1990) argues that problem framing has a large influence on the strategies used and roles played by non-governmental organizations. Those who define the communities' problem as shortage, e.g., shortage of food and services, tend to focus on relief and welfare (first generation NGOs); those that diagnose the problem as local inertia—that communities are not doing enough to better themselves—tend to focus on community development and self help strategies (second-generation NGOs). Those that diagnose institutional and policy constraints tend to focus on sustainable systems development; and those that diagnose inadequate mobilizing tend to focus on building coalitions and strengthening networks (Korten 1990).

House (1979) offers some diagnostic assumptions that underpin urban development. One assumption views the slum as a once healthy community that has become diseased. Stakeholders operating on such an assumption may envision wholesale redesign and reconstruction as the solution. Another view of the slum portrays it as a viable, low-income natural community which offers its residents important social benefits. The second view may imply strikingly different prescriptions for improving the community such as building the capacity of slum residents to meet specific needs like food, without necessarily taking them out of the slum.

Prescriptive assumptions are similar but different from transformational assumptions discussed in this chapter. The two assumption categories relate to how a program or intervention works to address the problem. The difference is that prescriptive assumptions are embedded in the decisions or choices of the most appropriate strategy or intervention for the problem. When this intervention is

selected, there are assumptions on how the intervention's immediate results (outputs) are expected to elicit long-term changes (outcomes or impact). As is the case with diagnostic assumptions, prescriptive assumptions are seldom made explicit. Occasionally, they are unveiled during program evaluations. Reading several evaluation reports, it is possible to note such assumptions unearthed. Sometimes the lack of clarity in stating these assumptions from the beginning has been given as the explanation for lack of program success. Is it too late to unearth prescriptive assumptions during evaluation? No. it is not too late. Unearthing prescriptive assumptions can help in clarifying the program's performance story (Mayne 2011). However, it is not timely enough.

Explicating Prescriptive Assumptions

Goal Evaluation

A technique that comes close to unearthing stakeholders' assumptions about the selected strategies or other possible strategies is "goal evaluation" or "engaging the goal paradox" suggested by Friedman et al. (2006) in their paper : "The Power of Why: Engaging the Goal Paradox in Program". In this case, the strategies are presented in the forms of goals. The approach is not based on the assumption that in most programs, differences among stakeholders require forging some degree of common identity. Program goal setting as identity formation involves asking three fundamental questions: a) What are your definitions of success, or goals, for this program?; b) Why are these goals important to you?; c) How do you think the program should go about achieving these goals? (Friedman et al. 2006). In the goal evaluation technique, each of these questions is asked on the individual level ("Who am I?"), the stakeholder group level ("Who are we?"), and the program or intergroup level ("Who are we as a coalition of groups?").

Rather than minimize or obscure differences, this approach aims to make differences at all levels as visible as possible so they can be seriously engaged and resolved before a program actually takes shape. Discussions focus on the why question: Why are these goals important to you? Why do you care passionately about them?

This discussion around goals provides a chance to unearth stakeholder assumptions about different goals and what they mean in context of a need or problem: "When we ask an individual why she or he feels passionate about his or her goals, we are not asking for explanations, justifications, or rationalizations. Rather, we want the individual to reflect on and articulate the underlying values, experiences, and contexts in which these goals are rooted. The aim of the why question is to enable different stakeholders to authentically express themselves so that they can discover deep commonalities and appreciate differences. This process

enables stakeholders to question their own goals, reframe them, and discover new ones" (Friedman et al. 2006, p. 204).

This is where assumptions can be unlocked—when stakeholders begin to describe their individual and groups understanding of or definition for success, for what works, or what sort of intervention out to be undertaken as opposed to the other. Since tacit assumptions are often based on such beliefs, an exercise like this one can go a long way in unearthing stakeholders' prescriptive assumptions.

A Strategy Assessment Approach

Another technique that comes close to unearthing tacit prescriptive assumptions is the strategic assessment approach (SAA) suggested by Leeuw (2003, pp. 10–13). This approach (also called "assumptional analysis") is based on dialogs, with four major stages:

(1) Group formation, where a wide cross-section of individuals with an interest in the relevant policy question are divided into groups.
(2) Assumption surfacing, where the different groups separately unearth the most significant assumptions that underpin their preferred policies and programs. Through the stakeholder analysis technique, each group is asked to identify the key individuals or groups upon whom the success or failure of their preferred strategy would depend (e.g., who is affected by the strategy? Who has an interest in it? Who can affect its adoption, execution, or implementation? And who cares about it?). For the stakeholders identified, each group then lists what assumptions it is making about each group in believing that its preferred strategy will succeed. Stakeholder analysis is followed by assumption rating. This involves use of secondary (program-related documents) and primary data (interviews or discussions with key stakeholders) to search statements about symptoms of the problem that have to be addressed by the strategy. These symptoms are distinguished from statements about causes of the problem. For each of the listed assumptions, each group asks itself (a) How important is this assumption in terms of its influence on the success or failure of the strategy? and (b) how certain are we that the assumption is justified? At this stage, the listed assumptions can be assessed. Each group then is able to identify a number of key assumptions (most important and least certain) upon which the success of its strategy rests.
(3) Dialectical debate, where the groups are brought back together and each group makes the best possible case to the others for its preferred strategy, while identifying its key assumptions. This is followed by a debate focusing on which assumptions are different between groups, which are rated differently, and which of the other groups' assumptions each group finds most troubling. Each group develops a full understanding of the preferred strategies of the others and their key assumptions.

(4) Synthesis, where assumptions are negotiated and modifications to key assumptions made. Assumptions on which the stakeholders (participants) agree are outlined. These may be the basis of a new strategy, for example, if stakeholders realize that the strategy has been based on faulty assumptions.

A key strength of the strategic assessment approach is the focus on group dynamics. Sharing knowledge and sharing perspectives when doing a reconstruction is central (Leeuw 2003). The approach has been criticized for not specifying any criteria for assessing the validity of the assumptions. Similar to the PSA, this approach has focused on existing policies, unearthing assumptions behind a strategy in view of the problem purportedly addresses. It can be even more helpful for stakeholders to unearth these assumptions before the strategy is implemented. This can save resources and time.

The Alternative Strategies Approach

Prescriptive assumptions relate to the assumptions about an intervention or strategy devised to address the problem or to reach a stated objective. Once the core problem is identified, these assumptions are converted into an objectives tree. The objectives tree can therefore be conceptualized as the positive mirror image of the problem tree, and the "cause and effect" relationships become "means to end" relationships. There may be gaps in the logic of the initial objective tree that were not apparent in the problem tree, therefore the 'means-ends' linkages between objectives should be reviewed and reorganized as necessary (See Fig. 9.1).

Finally, objectives dealing with similar topics can be grouped together in clusters, which will provide the basis for Strategy Analysis. Once complete, the objective tree provides a comprehensive picture of the future desired situation (see Fig. 9.2)

A review of the problem analysis may lead to the emergence of a different focal problem at a later stage, but this does not affect the validity of the analysis (Cockerill et al. 2000)

Once complete, the problem tree represents a comprehensive picture of the existing situation. The stage of the analysis phase involves the selection of the strategies which will be used to achieve the desired objectives.

Strategy analysis involves deciding what objectives will be included in the project, and what objectives will remain out, what the project's purpose is, and what the overall objectives will be. In addition to examining the logic, strategy analysis also looks at the feasibility of different interventions.

Depending on the scope and amount of work entailed, the selected clusters or strategy may form a project-sized intervention, or a program consisting of a number of projects. In the example above, our project will address the primary and secondary healthcare strategies, but not the nutritional awareness strategy, which is addressed by another project (EC 1999a) (see Fig. 9.3).
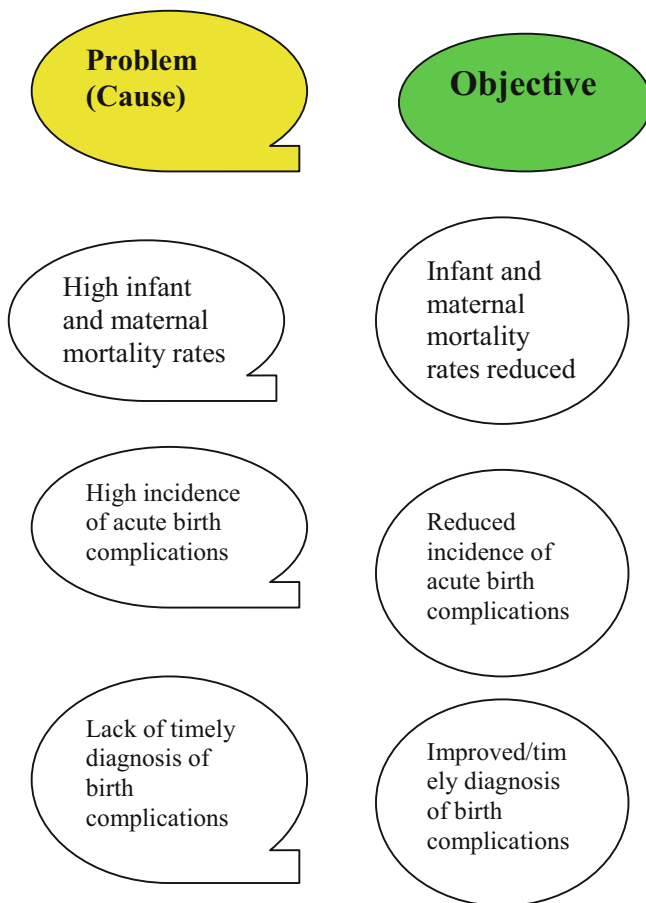
Fig. 9.1 Forming an objectives tree for a maternal and child health program (Adapted from EC 1999a)

Once the major objectives to include in the intervention have been determined, stakeholders have to identify which strategies to employ. There are a number of alternatives from which stakeholders can choose. The choice of an alternative or a combination of alternatives depends on the characteristics of those alternatives (see Table 9.1), such as:

(a) The alternative is efficacious and the best remedy of the problem, therefore it is the best way to achieve the objective. The strategy suits the context; hence it is contextually relevant and practical

(b) The alternative is feasible. We can implement this strategy within the project life and with the available resources, including human resources, finances, etc. For example, stakeholders may realize the need for an irrigation scheme to
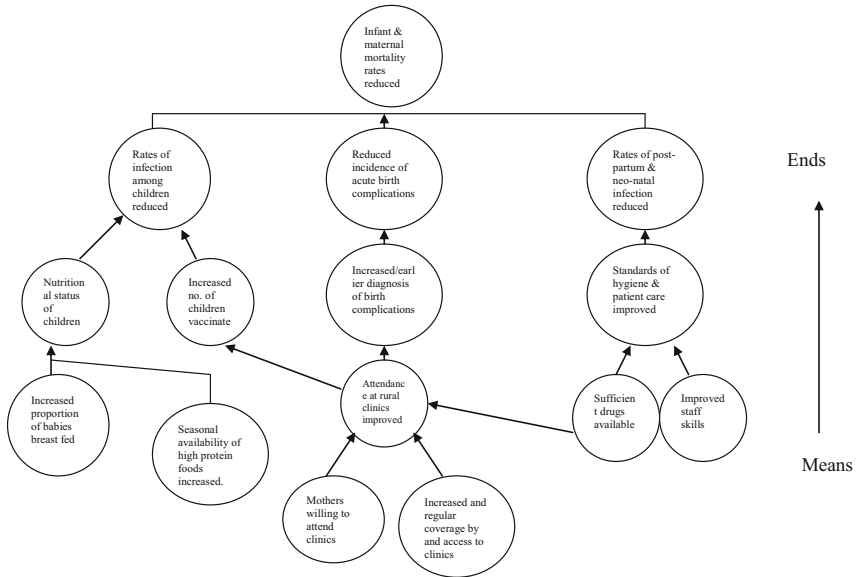
Fig. 9.2 An objectives tree for a maternal and child health program (Adapted from EC 1999b)

boost agriculture production, yet this may not fit within the program resources. Perhaps, they may resort to lobbying for government policy in this direction and or providing drought-resistant crop varieties.

(c) The alternative is the most sustainable solution to the problem, thus if the strategy is implemented, the root causes will be reversed and the new condition will be durable. Providing meals to malnourished children may reduce malnutrition significantly in the short run, but malnutrition may return soon after the program winds up.

(d) Where does the strategy fit with in everything else happening within this environment? What are other actors like government doing? Stakeholders may realize that they need to strengthen the health services delivery system, while the government might be in the process of building infrastructure and strengthening staff capacity; provision of tools and equipment might be what is missing.

Other criteria may come into play, but a decision on the most appropriate strategy will depend on collaboration embracing all these criteria.

The decision by stakeholders on which should be the most appropriate strategy to take could be facilitated by attaching scores to the various options based on a set of criteria, but the scores themselves may not be as important as how they are interpreted and the assumptions made in their interpretation.

Once the appropriate option is taken, the criteria for taking that option or set of options needs to be stated clearly in the context of efficacy, feasibility, relevance, sustainability, and fit with other interventions going on.
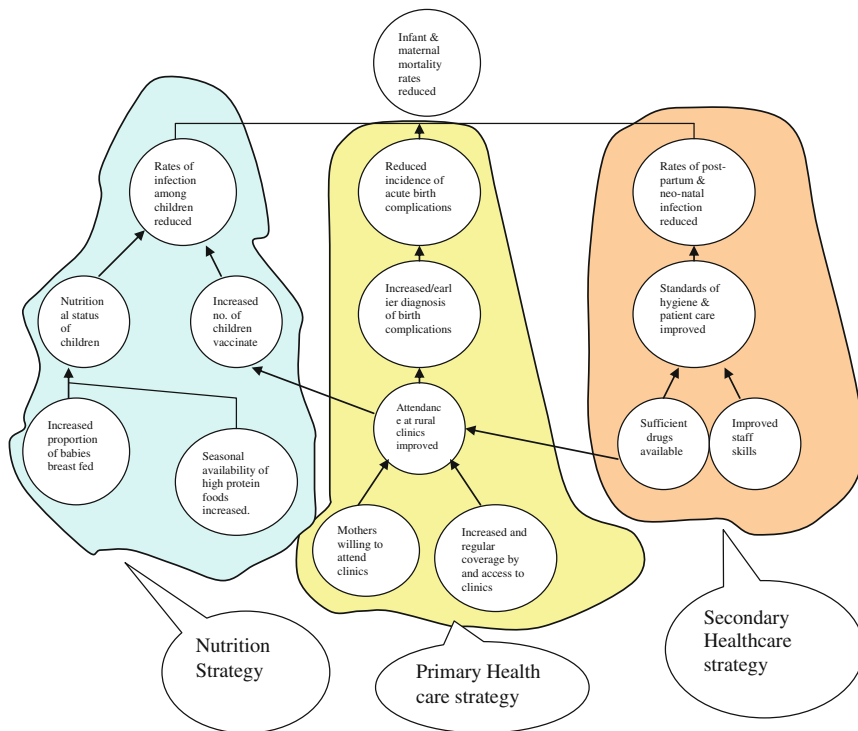
Fig. 9.3 Selection of strategies

For the example above, stakeholders may decide that to teach communities cost effective ways of treating water, as well as advocating that the government construct more clean water access points in the community are the most appropriate options.

These may have been selected for a number of reasons:

- Methods of water treatment that are available make the water safe enough for use; they are affordable and will easily be used consistently by communities.
- The considerations for eliminating other options relative to the selected option need to be stated. For example, water treatment may not be as new or foreign as bottled water in the community; hence individuals may be likely to use the bottled water alongside the locally available 'unsafe' water.

In stating the considerations on which an option is selected as well as the risks to those considerations, stakeholders would be stating the prescriptive assumptions.

Table 9.1 Alternative strategies matrix

	Option 1 (e.g.) Build more infrastructure (water points)	Option 2 Teach communities cost effective ways of treating water (the diagnostic assumption posits that people understand that they should drink 'safe' or boiled water)	Option 3 Supply bottled water to households	Option 4 Conduct an advocacy campaign for the government to build a large dam or more water points
Effectiveness	This option may drastically increase access to safe water and hopefully reduces the number of people who use unsafe water hence reducing the prevalence of water borne illnesses. In this case, this is based on a tested theory that having safe water is not enough if the safe water is not available	This alternative may help communities to make better use of the unsafe water available	This alternative will certainly give communities an alternative to the unsafe water	It may be long after the program life that the advocacy efforts could bare fruit depending the existing priorities of government
Contextual relevance	There are cultural beliefs on who should access or control the water points (gender) and this has to be taken into consideration in spelling out the strategy	The approaches of water treatment are subject to a range of contextual factors, e.g. perceptions about a chemical used to treat the water and how it could be potentially dangerous	The majority of community members are OK with using this water depending on how they perceive it…e.g. some keep it and sell it for food or others uses	Having government take action is well placed and consistent with the way business is done in this community, since government also provides other services like health

(continued)

Table 9.1 (continued)

	Option 1 (e.g.) Build more infrastructure (water points)	Option 2 Teach communities cost effective ways of treating water (the diagnostic assumption posits that people understand that they should drink 'safe' or boiled water)	Option 3 Supply bottled water to households	Option 4 Conduct an advocacy campaign for the government to build a large dam or more water points
Feasibility	The program may not have enough resources to construct 'enough' water points to create the effect required	Some water treatment options are affordable, especially if all it takes is to teach communities how to use the approach	The program might have a budget for a sizable proportion of the households during the program year	The advocacy activities are within in budget and capacity for the program
Sustainability	The water points will require maintenance even after the program winds out and this should be considered in spelling out the strategy	Some treatment methods may use chemicals or materials that are not locally produced, such wouldn't be sustainable	Community members will most probably not afford this water after the program stops providing it and they will return to the unsafe water thereafter	The government and the community will outlive the program in the same way, hence services provided by government are more sustainable than those provided by the program or NGO
Fit with other strategies being implemented	Nobody else is doing anything to improve water and sanitation in this community			

References

Brandenburg, R. (2009). Assumption interrogation: An insight into a self-study researcher's pedagogical frame. In D. L. Tidwell et al. (Eds.), *Research methods for the self-study of practice, self-study of teaching and teacher education practices* (pp. 195–210). New York: Springer.

Brookfield, S. D. (1995). *Becoming a critically reflective teacher*. San Francisco: Jossey-Bass.

Cockerill, R., Myesis, T., & Allman, D. (2000). Planning for Community-based Evaluation. American Journal of Evaluation *20*(3), 351–357

European Communities (EC). (1999a). *Project cycle management training handbook*. Brussels: EC.

European Commission. (1999b). *Joint relex service for the management of community aid to non-member countries (SCR) Resources, relations with the other institutions, evaluation, and information evaluation*. Sussex: EC.

Friedman, V. J., Rothman, J., & Withers, B. (2006). The power of why: engaging the goal paradox in program evaluation. *American Journal of Evaluation*, *27*(2), 201–218.

House, E. R. (1979). Coherence and credibility: the aesthetics of evaluation. *Educational Evaluation and Policy Analysis*, *1*(5), 5–17.

Korten, D. C. (1990). *Getting to the 21st century:Voluntary action and the global agenda*. West Hartford: Kumarian Press

Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation*, *24*(1), 5–20.

Mayne, J. (2011). Contribution analysis: addressing cause effect. In K. Forss, M. Marra, & R. Schwartz (Eds.), *Evaluating the complex: attribution, contribution, and beyond* (pp. 53–96). New Jersey: Transactional Publishers.

Myers, J. (1995). The value-laden assumptions of our interpretive practices. *Reading Research Quarterly*, *30*(3), 582–587.

World Vision International (2007). Learning through evaluation with accountability and planning: World vision's approach to design, monitoring and evaluation (2nd edn.). Monrovia: World Vision International.

Chapter 10
Transformational Assumptions

Abstract While prescriptive assumptions are related to strategies (and alternatives) devised to address a problem, transformational assumptions relate to how the immediate results of a strategy program or intervention (outputs) are expected to lead to long-term desired changes. The use of *Elicitation,* Program theory matrix and the theory of change Approach for *explicitizing* transformational assumptions is appraised. An integrative approach is proposed.

> There are certain propositions which philosophers, at one time
> or another, have said we 'unthinkingly assume' or
> 'instinctively believe'. To say that this is something we
> unthinkingly assume is to imply that it is a matter for dispute
> whether it is true. Talk of 'un-thinking assumptions' is thus tied
> up with the idea that the philosopher's job is somehow to
> justify, or else show to be un-justified, what we ordinarily
> neither question nor think of questioning
>
> Vesey 1954, p. 226

Introduction

Transformational assumptions are what Mayne (2011) refers to as 'behind the scenes' assumptions. Chen (2005) refers to them as a descriptive assumptions; Anderson (2004) refers to them as 'pre conditions' in the outcomes pathway which is an illustration of the relationship between actions and outcomes and also shows how outcomes are related to each other over the lifespan of the initiative. It is the most easily recognized component in a theory of change, because there are many planning approaches that employ boxes and arrows to depict program elements. Everything in the pathway of change is a precondition to the long-term goal. That is, the outcomes
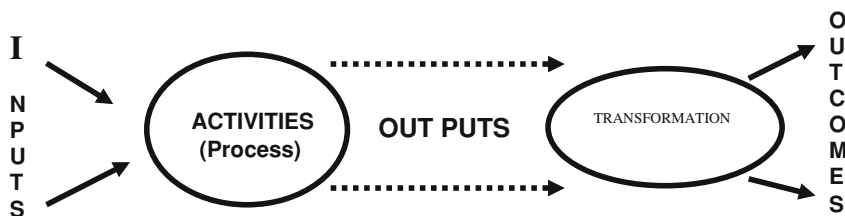
Fig. 10.1 Inputs, activities, outputs and out comes

of the path are all required to reach the goal—without each of them in place, we assume the goal cannot be attained. Assumptions explain both the connections between the preconditions for long-term change that occur in the early and intermediate stages of the change process, and the expectations about how and why proposed interventions will bring them about. According to Chen (2005), transformational assumptions are assumptions about a causal process through which an intervention is supposed to work—its success depends upon their validity.

A distinction between prescriptive and transformational assumptions has been made in the previous chapter. This distinction is worth reiterating at this point. While prescriptive assumptions are related to strategies (and alternatives) devised to address a problem, transformational assumptions relate to how the immediate results of a strategy program or intervention (outputs) are expected to lead to long-term desired changes. Stakeholders may determine that the major cause of malnutrition is the lack of proper feeding practices. This implies that other possible causes, such as shortage of food; illnesses, lack of access to safe water, etc., are excluded as major causes. This exclusion may be based on evidence or stakeholder experiences. Such arguments may or may not be made explicit. However, this exclusion may not be obviously apparent to stakeholders. Such factors are either dismissed or unspoken; without clear guidance, it is very unlikely that these considerations will be uttered or documented. In either case, whether alternative problem causes are taken for granted unconsciously or considered but not explicitly, they are tacit assumptions. Consider a strategy of educating community members about proper feeding practices. Between the problem being surfaced and the program being implemented, there are many causal links that are presumed.

They generally involve transformational assumptions.

What makes transformational assumptions crucial is the fact that program stakeholders may have no control over the desired long-term outcomes, though they may have control over the program outputs. The difference between outputs and outcomes is sometimes referred to as the frame of utilization, as presented in Fig. 10.1 (IFAD 2000)

A common adage points to the challenge of eliciting outcomes: 'You can lead a horse to water, but you can not make it drink' Patton, in his book 'Utilization-

focused evaluation' has used this adage to illustrate the dilemma of 'having no control over outcomes.' And he explains:

> The desired outcome is that the horse drinks the water. Longer term outcomes are that the horse stays healthy and works effectively. But because program staff know that they can't make the horse drink the water, they focus on things that they can control: leading the horse to the water, making sure the tank is full, monitoring the quality of water, and keeping the horse within drinking distance of the water. In short, they focus on the processes of water delivery rather than the outcome of water drunk. Because staff can control processes but cannot guarantee attaining of outcomes, government rules and regulations get written specifying exactly how to lead a horse to water. Quality awards are made for improving the path to water-and keeping the horse happy along the way. Most reporting systems focus on how many horses get led to the water, and how difficult it was to get them there, but never get around to finding out whether the horses drank the water and stayed healthy (Patton 1997, p. 157).

To take Patton's illustration a little further: what if it is later discovered that the horse is healthy? Would the conclusion be that it had been drinking the water? Or what if the horse gets sick? Would the conclusion be that it did not drink water? Perhaps. Patton's illustration highlights a challenge that is crucial-the tendency to measure what is easy to measure rather than what is most important to measure. But there is even a greater challenge: the fact that stakeholders' beliefs of how 'what is easy to measure' (taking the horse to the well) should translate to 'what is difficult to measure' (drinking and staying healthy) are often not clarified. It may be presumed that the more the horse is taken to the well, the more it drinks or the more water is provided, the more the drinking; or the more the horse drinks, the more likely it will be healthy; etc. Once such assumptions are made explicit, there is more chance that they will be examined or tested, and if they are faulty, new strategy will be formulated.

Transformational assumptions are not often spelled out in program designs. Sometimes they are revealed in evaluations. In reviewing many evaluation reports, it is possible to identify transformational assumptions that have been unearthed. As it turns out, these are sometimes said to be faulty, which is the reason the program was not successful. Below are some examples.

In Nelson (2011), the agriculture project was redesigned and integrated with health and nutrition, with the expectation that increasing agricultural production itself would not necessarily elicit improved nutrition for children. Integrating agricultural production with nutrition would ensure that communities make proper use of food to improve nutrition of their children, e.g., by educating them about appropriate feeding practices. This also assumed that the households targeted with agricultural production interventions were those that received nutrition education.

In Nelson (2011), it was determined that improving access to quality education would be achieved by providing functional literacy and life skills to children, increasing community involvement in education, advocating children's right to education, and improving the school environment. It could have been assumed that access to education is both a supply and demand problem; supply being the quality of education provided, demand being the existence of cultural beliefs that would

keep children away from school. Thus, improving quality would make it attractive for children to come to school in addition to eliciting better learning outcomes (demand is a function of quality); and at the same time communities would be more aware of the need to send children to school.

Project planners decided that improving the health and nutrition status of mothers would be achieved by increasing their access to health services, increasing prevention of infectious illnesses, improving their nutrition status, and improving the motivation of health workers. It could have been assumed that health status was mostly a supply problem—mother and children are suffering poor health because inadequate health services are provided; and not that poor health is a consequence of their behavior or that health services are provided with poor quality.

Stakeholders could argue that achieving food security would be attained by increasing food availability and household income and reducing post-harvest waste. It could have been assumed that prices would remain stable and favorable, such that increasing agricultural production would elicit proportionate increases in incomes. But there is risk in taking these assumptions for granted, a risk that food security would dissolve into food insecurity for a sizeable element of the population.

Making Transformational Assumptions Explicit

While assumptions are often the set of beliefs that guide a group (and often remain unstated), they may also be supported by research, or 'best practices,' which can strengthen the case to be made about the plausibility of theory and the likelihood that stated goals will be accomplished. Statements are made about how and why we expect a set of outcomes to come about as depicted in the pathway of change. These statements can reflect understandings of the change process taken from research, or they can be taken from practical experience. They should also reflect an understanding of the context within which a program operates. Often assumptions raise questions about the extent to which we can bring about the change we envisage.

Theoretical Frameworks: A Nook for Transformational Assumptions

A common form of articulation of descriptive theory, and thus a nook for transformational assumptions is theoretical frameworks (Chen 2006; Weiss 2000; Fishbein et al. 2001). Theoretical frameworks do not only relate to transformation theory and assumptions, many of them start with a diagnostic position: what

causes a problem, followed by the prescriptive and transformational positions: What causes the problem, what is the best strategy to address it and how does this strategy elicit intended outcomes?

A water and sanitation program could be based on one of two different theoretical frameworks; health based or integrated. A health-based approach presumes that poor health caused by poor water supply, quality, insufficient sanitation, and unsafe hygiene behavior is regarded as both a symptom and cause of poverty. The approach focuses mostly on developing government led water supply and improving sanitation (Nicol 2000). The 'integrated approach' argues that although supply of safe water is essential for health, it is only a part of the solution, as water can quickly become unsafe if people do not have adequate and appropriate sanitation facilities or the chance to develop good hygiene practices. This was modified into the demand-based approach, which argues that public sector delivery is financially unsustainable; and giving communities a financial stake in their own supply development is more likely to lead to durable systems (Nicol 2000). A program designed on these premises would have its outcomes pathway implied—that 'safe water' plus appropriate sanitation facilities and improved hygiene behavior will lead to improved health. In this pathway transformational assumptions are implied—for example, that individuals have control over their health, but not complete control.

Another commonly used theoretical framework is the sustainable livelihoods (SL) approach (e.g., Ashley and Carney 1999; Toufique 2001. Hans et al. 2006). The focus of 'livelihood' in SLframeworks is an attempt to move away from narrow definitions of poverty. As such, it reframes the broad aim of development as an effort to improve people's livelihood options. 'Livelihood' refers broadly to a means of making a living, and includes the assets, access to institutions and processes, and strategies that a person utilizes to achieve livelihood outcomes (Ashley and Carney 1999). The term 'sustainable' refers both to the characteristic of a livelihood to endure the various shocks and uncertainties likely to be encountered in the environment, and to avoid contributing to long-term depletion of natural resources.

There are five types of assets, pictured as corners of a pentagon, which are meant to emphasis their interrelatedness. The five assets are human, social, physical, natural, and financial. These can also be referred to as human capital, social capital, and so on. The terms 'asset' and 'capital' are roughly interchangeable in this framework then, except that the term 'capital' emphasizes the concept of potential investment or depletion (Pretty 2003). Each actor may either own or have access to a unique set of these assets, although patterns of ownership or accessibility may be discerned within a given social context.

The concept of structures and processes is quite broad and brings in all the social and cultural factors that may shape livelihoods. It is mainly at this point that the SL framework attempts to link macro and microfactors. These structures and processes may be accessible to people who may use them in their strategies, but they also shape and constrain people's possibilities.

People work within all of these factors—the risks they face, the assets they own or can access, and the social reality of their culture, system of government, and private sector, to weave livelihood strategies which result in livelihood outcomes. These outcomes can include increased income and savings, but more broadly, include improved well-being, reduced vulnerability to risk, and improved future livelihood options, and the sustenance or even replenishment, of natural resources.

This theoretical approach has been adapted to form program theory and to guide evaluations in many sectors. Parkinson and Ramirez (2001) used the framework to study the impact of ICTs in Colombia. In this study, the transformational assumption was that the availability of the Internet at a tele center would lead to improved livelihood outcomes for local residents, especially those who face the most limitations on their livelihood options.

Hans et al. (2006) used the framework to identify and test a number of transformational assumptions pertaining to rural development policies and sustainable land use in Honduras. Their study found that soil fertility had a strong, direct, and positive impact on income, while agro-climatic conditions such as higher rainfall and altitude had an indirect positive income effect, because they stimulated more remunerative livelihood strategies. They also found that land was not the key constraint limiting the potential for higher incomes. Land tenure had no impact on crop productivity and household income, but adoption of sustainable land use practices was higher on owner operated plots than on leased ones.

Consider an example from health communication. This sector is inundated with a variety of behavioral change communication theories. A common one is Bandura's social cognitive (learning) theory, which specifies that audience members identify with attractive characters in the mass media that demonstrate behavior, engage emotions, and facilitate mental rehearsal and modeling of new behavior. The behavior of models in the mass media also offers vicarious reinforcement to motivate audience members' adoption of the behavior (Kincaid et al. 2001). This implies that a health communication program based on this model will seek to pass on outcome-related messages though a character that is attractive to the target audience. In evaluating such a program, the frequency of receiving the message by the target audience may be as important as the frequency or intensity of identifying the message with the attractive character, in relation to internalizing the message, and adopting the desired behavior.

In the education sector, the 'whole school approach to school based quality reform' has been used to design, implement, and evaluate many programs (e.g., Lulua et al. 2007; Tucci 2009). According to this approach, improving teacher capacity, school management, community involvement in education, and learning environments altogether will lead to more effective institutional governance, which will elicit improved learning outcomes. In evaluating this approach Vernez et al. (2006) made the assumptions that learning outcomes would depend on the extent to which model elements have been implemented, how long it takes a core model element, such as improving teacher capacity, to be implemented and the extent to which external assistance to schools conformed to the model. To what extent does actual implementation conform to model, among others.

Theoretical frameworks do not necessarily articulate the transformational assumptions. The frameworks usually articulate a sequence of outcomes that may emerge from implementing a strategy that is based on an understanding of problem causes. The transformational assumptions are usually embedded or implied in the chain of outcomes and other concepts related to structural issues outlined by the model or theory. Sometimes the very principle that generates the sequence of outcomes also serves as the transformational assumption. In the water example…the recognition of individuals role in their health and their need for safe water options would underpin the combination of safe water provision and behavioral change strategies to reduce water-related illnesses in the community.

Theoretical frameworks may not necessarily be tested. Even when they are tested in one environment, they may not apply the same way in another. They have to be adapted to the program context and situation. The value in them is that, where they are appropriate, they provide the first step to coherent and clear thinking about the program and how it could or could not be implemented. The power of an explicit theory is that it gets to be known and subjected to logical and empirical testing. The challenge is that nonexplicit (implicit) theory at worst, cannot be known and at best, can only be guessed. Yet, this implicit theory could contribute a lot to designing, implementing, and evaluating a program. Chen (2005, p. 41) emphasizes the need to facilitate the explication of a program's implicit theory.

> Stakeholders have their own theory. Each of the stakeholders perceives the problem in a unique way and develops a personal preference for a particular intervention to address the problem. These perceptions and preferences may come from past experiences, conventional wisdom, discussion with peers, advice from experts, and scientific theories. Stakeholder theory is not usually as elaborate as scientific theory. It is implicit, yet so important as it influences stakeholder decisions on the course and process of implementing the program, and will influence their view of success of program (Chen 2005, p. 41).

Explicating Transformational Assumptions

The approaches discussed in this section are based on the assumption that an outcomes pathway or logic model results chain, or whatever you may call it, exists. The approaches focus on articulating the 'behind the scenes' assumptions.

The Elicitation Methodology

An approach that is close to unearthing transformational assumptions is the elicitation methodology. This approach is suggested by Leeuw (2003, pp. 14–15). Leeuw argues that policies and programs are developed and implemented by organizations, and that the 'mental models' or 'cognitive maps' of members of these organizations are important for understanding the anticipated impact of these

policies or programs. Thus, ideas about the change process—how program outputs will be changed or transformed into outcomes (long-term change) are contained in the cognitive maps or mental models of program stakeholders. It is the process of engaging stakeholders in articulating their mental models or cognitive maps that actually unearths these assumptions. Thus, elicitation is about involving and facilitating stakeholders in exercises during which they 'unpack' and verbalize their mental models/cognitive maps regarding how program outcomes will eventually be realized from outputs.

Steps in Elicitation

- Review the strategic intentions of the program or organization, especially through a study of the documentation which is designed to direct behavior;
- Obtain a deep understanding of how stakeholders make program decisions. Leeuw (2003) suggests that this could involve an anthropologic observation approach; watching decision makers and listening to their stories; or being involved in crisis moments of decision making;
- Conduct interviews and group discussions with stakeholders; discuss with them the observed decision-making situations to detect theories that guided their decisions and behavior. Such interviews need to be informal and interactive (Van der Heijden and Eden 1998). During the group discussion, participants should be encouraged to freely air their disagreements;
- Data generated are analyzed to generate conclusions on stakeholders' assumptions as verbalized in the discussions.

An Example of Elicitation

Johnson et al. (1998) and Leeuw (2003) describe an example of elicitation method in use. Twenty-two managers from international automotive industries were involved in discussing their mental models regarding the nature and causes of competition in the automotive industry. During this process, names of the automotive organizations including those of the participants' organizations, are written on small cards. These were then handed back to the managers who are asked to arrange them on the table in front of them to reflect the manner in which those organizations compete with each other. Participants were then asked to remap as many times as required. They were asked to explain the knowledge structures/maps they had just created. The justifications they verbalized were again written on small cards. This final set of cards, once given back to the managers and arranged, is considered to make up the construct map. The first map that the managers create of their competitive environment is considered the most salient and is taken forward for data analysis. Out of this exercise, 59 constructs on the

nature and causes of competition in the automotive industry were generated (Leeuw 2003).

Strengths and Weaknesses of Elicitation Methodology

A strength of this approach is to related stakeholder interviews or explanations of their mental models and assumptions in real crisis decision-making situations. This advantage is enhanced by the anthropologic observation approach to understand decision-making behavior, in order to adequately delve into their implicit theories and assumptions (Leeuw 2003).

However, elicitation has been criticized for lacking a knowledge base to guide the verbalization of mental models. Van der Heijden and Eden (1998) argue that it is not clear what to do with the assumptions once they are verbalized. In fact, some of the verbalized assumptions may be invalid; yet, verbalizing them poses a risk of perpetuating them even more. It is, therefore, important that elicitation of program assumptions is accompanied by clarification on the validity of assumptions to ensure that invalid assumptions are discarded.

Also, as explained by Leeuw (2003), Van der Heijden and Eden (1998) and others; elicitation has been used to (re) construct theories for existing or ongoing programs. This definitely improves the programs, chances of success, and its evaluability. However, it is necessary that stakeholders explicative transformational assumptions (why they think program outcomes will be realized) early enough, during program design. This timeliness ensures that the program won't be designed on basis of faulty assumptions. Also, as a program is implemented, monitored, and evaluated, the clarity of these transformational assumptions serves to remind stakeholders of why the program may or may not be performing as expected.

Using a Program Theory Matrix

One approach to bringing transformational assumptions to the fore is the use of a program theory matrix (PTM) provided by Funnel (2000). The PTM has seven components (see Table 10.1). The approach begins with outlining a sequenced hierarchy of intended outcomes. The hierarchy starts with program outputs such as the number of targeted beneficiaries reached by a program. This is followed by a series of intermediate outcomes, e.g., changes in knowledge and practices that, in turn, are followed by wider and long- term impacts (e.g., ultimately addressing the problem). For each identified outcome, a series of questions is raised, the answers to which are recorded in a matrix: What would success look like: for example, the nature of desired changes in knowledge with whom? What are the factors that influence the achievement of each outcome? Which of these can be influenced by

the program (e.g., quality of services delivery?) which factors are outside the direct influence of the program (e.g., economic climate)? What performance information is needed to be collected? How can stakeholders collect the information? (Funnel 2000, p. 92).

An Integrative Approach

Approaches like elicitation, and the PTM discussed earlier have been used mostly in (re)constructing theories for existing programs. The integrative approach discussed in this section assumes that transformational assumptions are explicated during the design phase of a program's cycle before it is implemented or evaluated.

Three approaches which involve elicitation of transformational assumptions are reviewed:

- The W. K. Kellogg Foundation's approach.
- The Aspen Institute Roundtable on Comprehensive Community Initiatives approach to explicating transformational assumptions.
- Mayne's contribution analysis.

The first two are commonly used during program design. Although contribution analysis is commonly used during program evaluation, it brings very useful features to this integrative approach. In integrating these three approaches, the following questions are used:

(a) Where in the program theory is the assumption focused?
(b) Does the approach define what the assumptions should look like?
(c) Does the statement of the assumption include the strengths of the assumption or the basis on which we can still hold it as a logical assumption?
(d) Does the approach specify the statement of the risks to the assumption?
(e) Does the approach guide stakeholders on thinking through the assumptions? (e.g., workshop tools)
(f) Context in which assumptions are stated.

The W. K. Kellogg Foundation's Approach

According to the Kellogg Foundation's (WKKF) (2004a, b) manual, assumptions are a crucial component of a program's theory of change (see Box 10.1 for a list of six steps to developing a theory of change). Assumptions are defined as statements of how and why the identified change strategies will work in the community (e.g., principles, beliefs, ideas).

Table 10.1 An example of application of program theory matrix to one level of the outcomes hierarchy for a small businesses pollution program

1. Intended outcome	2. Success criteria	3. Program factors affecting success	4. Nonprogram factors affecting success	5. Activities and resources of the program	6. Performance information	7. Sources of data
Change attitudes of target businesses toward being willing to change practices	Agreement by businesses to meet with program advisers with the view of identifying possible solutions; few refusals, preparation of action plans that include defined key elements and business specific examples of increased willingness	Availability of confidential credible advisory assistance Extent to which program can convince businesses of benefits of change Extent to which the program possesses are burdensome for businesses Success with which the program engenders industry association support	Extent of target businesses' 'illegal' polluting practices and whether businesses are prepared to take the risk of exposing their problems to the program in order to receive assistance. Businesses beliefs and past experiences concerning costs and benefits of change Whether the views and actions of competing businesses favor reducing pollution Size and nature of business and capacity to commit resources to preparing plans Whether the offer of assistance comes at the 'right time'	Promotes advisers and makes commitments about confidentiality Identifies and promotes proven benefits to similar businesses-uses case examples Offers and undertakes follow-up advisory visits Develops partnerships with industry associations Tailors paperwork and expectations to the type of businesses involved	Percentage of businesses that request and receive advisory assistance compared with targets set for each industry Percentage of businesses receiving advice that rate the advice as credible and useful (compared across target industries, and whether there is any relationship between perceived usefulness and apparent impact on willingness Percent of businesses that prepare action plans within the time frame of the program, compared with target and planning guidelines Percent of businesses that cite the advisory program as having contributed to willingness Examples of impact on willingness Reasons for nonparticipation or nonpreparation of plans and extent to which reasons relate to program factors and activities or nonprogram factors Extent to which successful partnerships are developed with the industry associations	Administrative records of requests for assistance and receipt of assistance Review of educational materials and strategies developed for the program and for each industry Post-program anonymous survey of business advisors Anonymous survey of businesses Site visits and case studies by evaluator Structured records of observations and notes kept by advisors Anonymous survey of businesses Interviews with advisors and key informants in industry associations Review of actions taken by industry associations

Box 10.1: WKKF's Steps in designing a theory of change

(1) Problem or Issue Statement
   Describe the problem(s) your program is attempting to solve or the issue(s) your program will address.
(2) Community Needs/Assets
   Specify the needs and/or assets of your community that led your program to address the problem(s) or issue(s).
(3) Desired Results (Outputs, Outcomes, and Impacts)
   Identify your desired results, or vision of the future, by describing what you expect to achieve, near- or long term, if your program is funded.
(4) Influential Factors
   List the factors (e.g., protective or risk factors, existing policy environment, or other factors) you believe will influence change in your community.
(5) Strategies
   List general, successful strategies or 'best practices' your research identified that have helped communities like yours achieve the kinds of results your program promises.
(6) Assumptions
   State the assumptions behind how and why the identified change strategies will work in your community (e.g., principles, beliefs, and ideas).

Source WKKF (2004a, b, p. 69)

After justifying the selection of a specific strategy from among the alternatives that have been researched, stakeholders have to state why this strategy is needed and why it will work in the community. It is important early on to document instances that describe the general condition of public reaction to the problem/issue and possible solutions.

Stakeholders should draw direct conclusions about the statement of need and capacities in the community in the assumption. In addition, it should be clear how the program intends to function as an intervention—to solve identified problems or build existing assets. Assumptions are listed last in the theory of change exercise because by this time, most of the information that supports assumptions should already be on the table.

An example is provided by the Kellogg Foundation Manual (2004a, b, p. 70)

As proven in Another town, access to affordable medical care reduces the incidence of emergency visits by providing appropriate, preventive primary care. A free medical clinic should prove successful in my town, because of its history of extraordinary volunteerism. My town's Medical Society officially encourages its 400 medical professional members to volunteer 20 h each year to help medically underserved community residents. My town's Nursing Association is also interested in collaborating with a free clinic. Memorial

Hospital has agreed to assist in planning and funding a free clinic. There is precedence for lending free facilities to medical projects serving those in need. My town's technical college donates space for my town's volunteer dental clinic. My town's Free Clinic will be strongly supported by the people, businesses, and institutions of my town, USA.

This technique of explicating assumptions could generate external, diagnostic, prognostic, and transformational assumptions alike. In justifying or stating the basis for the assumption, this technique also tries to ensure that there are minimal risks to the assumptions.

The Aspen Institute Roundtable on Comprehensive Community Initiatives (RTCCI) Approach to Explicating Transformational Assumptions

RTCCI's theory of change approach has been discussed in Chap. 2. This approach requires stakeholders to state assumptions about why each precondition is necessary to achieve the results along the pathway of change and why the set of preconditions is sufficient to bring about the long-term outcome (Anderson 2004). These assumptions may or should come from social science theory that connects program activities to specific outcomes for specific populations or communities. This may include findings from 'best practice' research as well as evidence from academic (or basic) research (Anderson 2004).

RTCCI suggests a number of questions that could help stakeholders to come up with assumptions: These include:

- When you look at the total picture, do you believe that the theory makes sense?
- Do the preconditions make sense as the logical steps toward the long-term outcome?
- How will we be able to bring about the outcomes at the levels we have predicted?
- Is there anything going on in the real world that may make it difficult to get this theory off the ground the way we have planned it?
- Is this theory of change PLAUSIBLE? Have we created a compelling story about the pathway of change that would lead to the long-term goal in this community?
- Is this theory of change FEASIBLE? Do we have the capacities and resources to implement the strategies that would be required to produce the outcomes in the pathway of change?
- Is this theory TESTABLE? Have we specified how success will be measured clearly enough that we can recognize progress toward our goal when we see it? Have we defined indicators for each outcome in clear terms that a researcher or evaluator can use to produce a research plan?

- These questions may not guarantee a definite pattern of assumptions, but they will definitely generate useful reflection on existing transformational assumptions.

Mayne's Contribution Analysis

We have discussed Mayne's (2011) contribution analysis in the previous section. Fear (2007) in Mayne (2011, p. 515) argues that program theory comprises a set of assumptions, beliefs, and (political) ideologies woven together to produce a formula and a statement of intent, to solve a particular problem. Thus, assumptions (in this case the transformational ones) are a crucial part of the program's theory. A program theory is incomplete without assumptions.

Mayne gives an example of a theory of change for enhancing planning, monitoring, and evaluation (PM&E) capacity in agricultural research organizations (AROs) adapted from Horton, Mackay, Anderson and Dupleich (2000). In this example, both the assumptions made about the change from one result to another and the risks to those assumptions are stated.

Integrating the Approaches

The three approaches to stating assumptions that we have reviewed can fit with each other. They all are similar in emphasizing that assumptions are a key component of a program's theory. Some of them emphasize elements different from others are listed in Table 10.2 below.

User Discretion is Advised

This book has reviewed approaches to stating assumptions (primarily transformational assumptions) that are presented as steps in broader processes. The Kellogg Foundation approach is part of a six-step process. The Aspen Institute's RTCCI approach is but a part of a five-step process. Mayne's approach is one step of seven in contribution analysis. The 'mosaic' approach we recommend neither disregards nor necessitates the inclusion of all other steps in any of these approaches. The use of the 'mosaic' approach assumes that the program theory is in place. It helps stakeholders to transform this theory 'results chain' to a more complete theory of change by articulating the transformational assumptions, the bases of and risks to those assumptions.

Table 10.2 Complementarities or differences among KWKF, RTCCI and Mayne's approaches to explicating assumptions

		KWKF's approach	RTCCI's approach	Mayne's contribution analysis
1	Where in the program theory is the assumption focused?	Generally focuses on why the program is expected to work- which could include external, diagnostic, prognostic, transformationa,l and other assumptions	Mostly focuses on assumptions between each set of preconditions in the pathway (transformational assumptions), but the questions provide room for other forms of assumptions	Mostly focuses on the joints among the results chains
2	Does the approach define what the assumptions should look like?	Yes: assumptions are stated as positive statements consideration that the necessary or anticipated factor or condition exists	Yes: assumptions are stated as positive statements consideration that the necessary or anticipated factor or condition exists	Yes: assumptions are stated as positive statements consideration that the necessary or anticipated factor or condition exists
3	Does the statement of the assumption include the strengths assumption or the basis on which we can still hold it as a logical assumption?	Requires stakeholders to justify/ rationalize or give a reason for believing that an assumption is worth while	Not explicitly required but the approach does not necessarily limit stakeholders from rationalizing assumptions	Not explicitly required but the approach does not necessarily limit stakeholders from rationalizing assumptions (this quality of an assumption may be attained in stating the risk to the assumption)
4	Does the approach specify the statement of the risks to the assumption?	Not explicitly required but the approach but this quality of an assumption may be attained in rationalizing the assumption)	Not explicitly required but the approach does not necessarily limit stakeholders from rationalizing assumptions	Yes: the approach specifies both assumption and risks to the assumptions
5	Does the approach guide stakeholders on thinking through the assumptions? (e.g. workshop tools)	No specific questions beyond 'why the program is expected to work'	Several questions are provided on why	No specific questions are provided beyond stating what a 'behind the scene' assumption is
6	Context in which assumptions is stated	Assumptions are stated in developing a theory of change	Assumptions are stated in developing a theory of change	Assumptions are stated in developing a postulated theory of change, during evaluation, to clarify the program's performance story

Focus Within the Program's Theory

As RTCCI and Mayne argue, transformational assumptions should focus on the 'joints or linkages' within the results chains. We have discussed other categories of assumptions in preceding sections of this book, particularly in the past two chapters, including diagnostic and prognostic assumptions that are described somewhat earlier.

What Should the Assumptions Look Like?

As depicted by all three approaches, assumptions should be stated as positive statements with consideration to the necessary or anticipated factors or conditions.

Stating the Strengths of the Assumption

As depicted in the Kellogg approach, stakeholders need to justify/rationalize or give a reason for believing that an assumption is worth while/logical

Stating the Risks to the Assumption

As depicted in Mayne's approach, it is important to state not only the strength, but also the risk in an assumption

Questions to Guide Stakeholders in Thinking Through Transformational Assumptions

If you have facilitated stakeholders through any form of planning, you already know that this is abstract and painful. You might ask stakeholders: 'What makes you expect that this program will work?' only to find they keep looking at you in surprise. It helps to have some questions prepared, even if they may not directly lead to the needed answers, As long as they generate good discussion, answers will eventually emerge. Using RTCCI's questions (or appropriate adaptations) can be extremely helpful in expanding thinking and identifying opportunities. The facilitator needs to use these judiciously and keep careful watch on the flow of the discussion, even introducing different questions.

Context in Which Assumptions are Stated

Assumptions need to be revealed, to whatever extent is possible, during the planning or program design phase. If these assumptions have never been specified, it is never too late to engage a reflection by stakeholders on these assumptions. As Mayne (2011) argues, getting to these assumptions will deepen the understanding of what evidence or data collected by the programs' monitoring and evaluation processes really means in terms of the program's performance and actions needed to improve it. However, sooner is better for informing program design and enriching evaluation strategies.

For example, if a program intended to reduce malnutrition has a theory that increasing agricultural production at a household level would ensure that children have enough food to eatand therefore they would be better nourished. If the reflection on assumptions only emerged during evaluation, and reveals the originally implicit assumptions such as:

- Increased amounts of food will not be sold off to the market to the extent that leaves children without food.
- Incomes from food sold will be used to feed children better.
- More food at the household level will necessarily mean that the children have balanced diets.
- Caregiver behavior will be in keeping with hygiene standards.

If such assumptions were generated during evaluation, the evaluator will know where to find helpful information on the programs performance. For example, assessing levels of malnutrition in the target population before and after the program would not be helpful without comparing the data with these assumptions as confounders. Or if the data show that malnutrition decreased, should the program stakeholders rejoice? (For malnutrition to reduce is not bad news after all). If the data show that malnutrition did not decrease, should they conclude that the program did not work? With in such a situation, the questions on the program's performance can not be answered effectively without getting 'behind the scene' assumptions into the picture. Unfortunately, it is not uncommon for programs to be evaluated with no reference to assumptions behind their theory (some programs may not even have an articulated theory). It could only help to have program theory and assumptions articulated as soon as possible, while lessons could be learned and courses altered. Having all this in place could only help to strengthen the evaluation and provide a growing body of knowledge to make these programs as good as possible. The evaluator, but mostly the stakeholders would be so lucky!

References

Anderson, A. (2004). *Theory of Change as a tool for strategic planning: A report on early experiences*. Washington, DC: Aspen Institute.

Ashley, C., & Carney, D. (1999). *Sustainable livelihoods: Lessons from early experience*. London: Department for International Development.

Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Newbury Park, California: Sage Publications.

Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools (Mid-South Educational Research Association)*, *13*(1), 75–83.

Hans, G. P., Pender, J. J., Damon, A., & Schipper, R. (2006). *Rural development policies and sustainable land use in the hillside areas of Honduras: A quantitative livelihoods approach, Research Report 147*. Washington, DC: International Food Policy Research Institute.

International Fund for Agricultural Development (IFAD) (2000). *Guide for M&E of rural agricultural projects.* Retrieved December 10, 2006, from http://www.ifad.org/

Fishbein, M., Hennessy, M., Kamb, M., Bolan, G. A., Hoxworth, T., Iatesta, M., et al. (2001). Using intervention theory to model factors influencing behavior change. *Evaluation and the Health Professions, 24*(4), 363–384.

Funnel, S. C. (2000). Developing and using a program theory matrix for program evaluation and performance monitoring. *New Directions for Evaluation, 87*(Fall), 91–101.

Leeuw, F. L. (2003). Reconstructing program theories: methods available and problems to be solved. *American Journal of Evaluation, 24*(1), 5–20.

Lulua, R., Aanyu, D., Sherburne, L. & Nkwake, A. (2007). *A Whole-School Approach to School-Based Quality Reform: The Uganda Experience*. Paper presented in the 51st Annual International Conference of the Comparative and International Education Society, Feb 25–March 1, Baltimore, USA.

Kellogg Foundation (2004a). *Using Logic Models to Bring Together Planning, Evaluation, and Action.* Battle Creek, Michigan: W.K. Kellogg Foundation.

Kellogg Foundation (2004b). *Logic Model development guide to help people help themselves through the practical application of knowledge and resources to improve their quality of life and that of future generations*. Battle Creek, Michigan: W.K. Kellogg Foundation.

Kincaid, D. L., Figueroa, M. E., Storey, D., & Underwood, C. (2001). *Communication and behavior change: The role of ideation*. Baltimore: Johns Hopkins University Bloomberg, School of Public Health, Center for Communication Programs.

Mayne, J. (2011). Contribution analysis: Addressing cause effect. In K. Forss, M. Marra & R. Schwartz (Eds.). *Evaluating the complex: Attribution, contribution, and beyond* (pp. 53–96). New Brunswick, New Jersey: Transactional Publishers.

Nelson, C. (2011). *Evaluation of Lugbu Area Development Program*, World Vision Sierra Leone. Unpublished Report.

Nicol, A. (2000). *Adopting a sustainable livelihoods approach to water projects: Implications for policy and practice*. London: Overseas Development Institute.

Parkinson, S. & Ramírez, R. (2001). Using a sustainable livelihoods approach to assessing the impact of ICTs in development. *Community informatics 2*(3). Retrieved March 13, 2012, from http://ci-journal.net/index.php/ciej/article/view/310/269

Patton, M. (1997). *Utilization focused evaluation: The new century text*. New York: Sage.

Pretty, J. (2003). *Social capital and connectedness: Issues and implications for agriculture, rural development and natural resource management in ACP countries.* Netherlands: Technical Centre for Agricultural and Rural Co-operation (CTA). Retrieved February 13, 2012, from http://internships.cta.int/pubs/wd8032/WD8032.pdf

Toufique, K. (2001). *Rights and livelihoods*. Workshop, Dhaka, Bangladesh: Bangladesh Institute of Development Studies and Power and Participation Research Centre.

Tucci, T. N. (2009). *Whole-School Reform: Transforming the Nation's Low-Performing High Schools, July 2009, Policy Brief, Alliance for excellent education*. Retrieved November 11, 2011, from http://www.all4ed.org/files/PolicyBriefWholeSchoolReform.pdf

Van der Heijden, K., & Eden, C. (1998). The theory and praxis of reflective learning in strategy making. In C. Eden & J.-C. Spender (Eds.), *Managerial and organizational cognition: Theory, methods and research* (pp. 58–76). London: Sage.

Vernez, G., Karam, R., Mariano, L. T. & DeMartini, C. (2006). *Evaluating Comprehensive School Reform Models at Scale Focus on Implementation*. Santa Monica: RAND Corporation.

Vesey, G. N. A. (1954). Unthinking assumptions and their justification. *Mind, New Series, 63*(250), 226–233.

Weiss, C. (2000). Which links in which theories shall we evaluate? *New Directions for Evaluation, 87*(Fall), 35–45.

Part IV
Conclusions

Chapter 11
Evaluating Assumptions

Abstract Examining the assumptions that hold a program theory together is a vital part of evaluating program outcomes. Examining implicit or explicit program assumptions facilitates understanding of program results-both intended and unintended. Evaluation approaches for testing program assumptions are outlined. The best place to start integrating assumptions in an evaluation is at the conceptualizing stage, when evaluation questions are being formulated, not in data collection or methodologies. Tools are but a servant of methods and methods a servant of questions, which should be the servant of objectives and/or purpose. Ideally, by framing the questions well, the methods, tools, and data will produce highly useful answers and solutions. But examining assumptions is without doubt a necessary element in the process.

Introduction

Davies (2010) in Shaw and Crompton (2003, p. 192) argues that for a theory to contribute to understanding, and to be of use, it must contain the following features:

- It must present a model that explains, integrates, and interprets facts.
- It must necessarily be prescriptive as well as descriptive, leading to conclusions about the proper or effective course for public policy
- It will be normative insofar as it will raise moral and philosophical issues and must adopt clear positions on these. Thus, meaningful and useful theory must contain, for example, relevant notions of the nature of 'the good society', the nature and causes of human action and the differentiation of public and private in the context of the project.

- Finally any effective theory must be historically aware—institutions and activities can only be properly understood in a historical context.

The above attributes imply that a theory must be as complete as possible. Perhaps the best example of a fuzzy theory is one that does not articulate assumptions that hold it together. Thus, these attributes cannot be required of a theory any more than they are required of assumptions, since assumptions are the glue that holds a theory together as a coherent concept. Assumptions are just a piece (yet a critical one) of the work but they help articulate, validate, and assess the theory and the project or program design.

Let me take you back to that story in the introduction: the expectant woman and the little girl. They were both involved in unfortunate disagreement because they each held different assumptions about what causes a big belly! You may realize that the differences in assumptions were not the core problem here. The problem was that these assumptions were not expressed. Neither of them imagined that the other might hold a different assumption. We are pretty sure that if the expectant mother had known why the little girl commented that way, the context for the entire discussion would have surely changed.

Framing the Questions

The best place to start integrating assumptions in an evaluation is conceptualizing, when evaluation questions are being formulated, not in data collection tools or methodologies. Tools are but a servant of methods and methods a servant of questions, which we hope, are the servant of objectives and/or purpose. Ideally, what is put into the purpose will get into the questions and all the way to the methods, tools, and data.

Thus the focus on assumptions is not isolated from evaluation questions. Assumptions are not necessarily the questions, but they may become the questions depending on the purpose of the evaluation. For example, to know that equipping teachers with better skills and materials contributed to better student academic grades may focus on measuring if the teacher skill improved, materials were used by teachers, and students' grades are improving. But evaluation would not focus on these questions any more than it would validate the assumption that poor grades are a result of bad teachers rather than poor student or unsupportive parents, or other factors. Conclusions of the evaluations need to be interpreted in light of these assumptions, even when the data may not be sufficient to test them.

As earlier mentioned, why we do evaluations influences how we do them. Assumptions are supposed to strengthen the role of theory in evaluations, and they might find a natural home in theory-based evaluation. But this does not mean that other evaluation approaches are not affected by program assumptions. McLaughlin and Jordan (1999) argue that there are essentially three purposes for evaluation:

a. to measure program performance (which they refer to as accountability)
b. communicating the value of the program to others and
c. program improvement

Green and McAllister (1998, p. 4) argue that despite the process used to develop the theory, the end product should answer the following questions: (a) what are the important program goals? (b) What are the program services offered? (c) What are the implicit or explicit assumptions about how program services will lead to program outcomes? This last question is key to developing a program theory, as it addresses the question of why and how the services provided are expected to lead to change in well-being of beneficiaries. Because of this focus, theory-based evaluation seems to lend itself more to the performance improvement purpose, which, according to McLaughlin and Jordan (1999), is in many cases the concern of stakeholders:

> When most managers are faced with accountability requirements they focus on collecting information or evidence of their programs' accomplishments—the value added for their customers and the degree to which targeted problems have been solved (Another way to be accountable is to be a good manager). Good managers collect the kind of information that enables them to understand how well their program is working. In order to acquire such an understanding, we believe that, in addition to collecting outcome information, the program manager has to collect information that provides a balanced picture of the health of the program. When managers adopt the program improvement orientation to measurement they will be able to provide accountability information to stakeholders as well as make decisions regarding needed improvements to improve the quality of the programs (McLaughlin and Jordan 1999, 71).

A common form of theory-based evaluation has come to be known as the theory of change approach to evaluation (see Prudence 1995; Green and McAllister 1998; Connell and Kubisch 1998; Tilley 2004; Scriven 2008; Chen 2006; Donaldson and Gooler 2002). This approach is said to be most suited for complex community programs (sometimes referred to as comprehensive community initiatives) that are left with a limited number of options for evaluation including: (a) retreating to process documentation of the initiatives and reducing expectations about obtaining credible evidence of their impacts; (b) try to "force fit" the initiatives themselves into the procrustean bed of existing and accepted evaluation methods in order to estimate their impacts; and (c) put off evaluating CCIs until the initiatives are more "mature" and "ready" to be evaluated using existing strategies (Connell and Kubisch 1998, 1). Weiss (1997b) agrees with these observations by arguing that some of the problems that beset theory-based evaluations include: inadequate theories about pathways to desired outcomes in many program areas, confusion between theories of implementation and theories of programmatic action, difficulties in eliciting or constructing usable theories, measurement error, and complexities in analysis. Connell and Kubisch (1998) suggest three steps to the theory of change approach to evaluating comprehensive community initiatives:

a. Surfacing and articulating a theory of change (help to specify the theory underlying the intervention and thereby "unpack" the intervention itself). This process is the one of conceptualization, where, first of all, the evaluation questions are clarified. Green and McAllister (1998) argue that one of the most important functions of a theory-based evaluation is to provide a basis for generating and prioritizing evaluation questions. The questions generated are meant to test the theory in three ways:

   i. Was the theory implemented as designed, and does that potentially affect its ability to deliver expected outcomes.
   ii. Are the mini-steps-outputs, short-term and intermediate outcomes visible.
   iii. Do the assumptions that link these mini-steps (including diagnostic, prognostic, transformational assumptions, among others), together hold? This way, a clear articulation of the theory and assumptions would be useful in framing the research questions.
      At the same time, the nature of questions will dictate the appropriate methodological designs needed to answer these questions. Theory-based evaluations have sometimes been blended in participatory and qualitative studies, and sometimes in mixed method studies as well (see Chen 2005, 2006). Take an example of the nutrition program. To measure the changes in malnutrition rates may take a quantitative approach, yet to understand the dynamics of how food at household level may or may not translate into improved child nutrition could require a qualitative approach. The bottom line is the methods should serve the purpose—to answer the questions.

b. Measuring a CCI's activities and intended outcomes. This involves translating the questions into observable measures to use in data collection.

c. Analyzing and interpreting the results of an evaluation, including their implications for adjusting the initiative's theory of change and its allocation of resources (tests of links between an initiative's activities and their expected outcomes). According to Weiss (1997a, b, p. 43) "If the evaluation can show the series of micro-steps that lead from inputs to outcomes, then causal attribution for all practical purposes seems to be within reach." This implies that analysis of data involves a triangulation of many pieces of information that 'draw a dotted line' or no dotted line at all between the intervention and the outcomes measured, by validating assumptions behind these links and reviewing other factors that would influence outcomes. This is what Mayne (2011) refers to as a program's 'contribution story'. This nature of triangulation is supported by Pawson and Tilley's 'realistic evaluation'. The characteristic of this approach is to stress what the components of a good program theory should be: context (C) and mechanism (M), which account for outcome (O). Evaluation is based on the CMO configuration. Programs are seen as the opportunities that an agent, situated inside structures and organizations, can choose to take, and the outcomes will depend on how the mechanism that is supposed to be at work will be engaged in a given context.

It is worth noting that there are several other approaches of theory-based evaluation. Some of these emphasize inbuilt performance monitoring that measures process and outcome indicators regularly, as opposed to summative evaluation. These approaches might have a combination of formative and summative evaluation elements, such as in Patton's developmental evaluation (say more on developmental evaluation). Other approaches have a combination of both formative and summative evaluation elements. For example, Schalock and Bonham (2003) recommend an outcome-based evaluation model that focuses on four categories of outcomes: organizational performance (outcomes), organizational value (process), individual performance (outcomes) (Schalock and Bonham 2003, 230)

This model of evaluation focuses on performance improvement. McLaughlin and Jordan (1999, p. 71) have elaborated this eloquently:

> Measurement strategies should involve ongoing monitoring of what happened in the essential features of the program performance story and evaluation to assess their presumed causal linkage or relationships, including the hypothesized influences of external factors. …the measurement should track the steps of the program. In the Logic Model the boxes are the steps that can often be simply counted or monitored\ and the lines connecting the boxes are the hypothesized linkages or causal relationships that require in-depth study to determine and explain what happened. It is the measurement of the linkages, the arrows in the logic chart which allows the manager to determine if the program is working (Monitoring the degree to which elements are in place even the intended and unintended outcomes will not explain the measurement or tell the manager if the program is working. What is essential is the testing of the program hypotheses (McLaughlin and Jordan 1999, p. 71).

This form of performance monitoring measures the theory progressively—as the short and intermediate outcomes are measured, the assumptions that connect these outcomes are also continuously tested, and necessary collective decisions can be taken to improve the program. For example, if a program aimed at reducing malnutrition through increasing food production and feeding practices stakeholders, may measure a series of mini-steps, including increases in land cultivated, yield per household, knowledge of appropriate feeding practices, and practice of appropriate feeding before it is time to measure changes in malnutrition. The program may not necessarily focus on improving hygiene and sanitation, but recognize that appropriate hygiene and sanitation practices are crucial in improving child nutrition. The process of monitoring feeding practices may as well include monitoring the existence of recommended hygiene and sanitation practices to determine if messages on feeding need to include hygiene or not.

Assumptions are Crucial in Evaluating Program Theory

Stakeholders need to evaluate program effects (outcomes and impacts). Evaluation may also be intended to and they always do. According to Weiss (1993), the evaluator who limits his study to program effects conveys the message that other elements in

the situation are either unimportant or that they are fixed and unchangeable. The intervention strategy is viewed as the key element, and all other conditions that may give rise to, sustain, or alter the problem are brushed aside. Several other program elements are normally evaluated—such as program process. Evaluation frameworks that simply measure program outputs and outcomes have a failure to link performance indicators with a strong evaluative criteria and standards (Funnel 2000).

Another important program element that is worth monitoring and evaluating is the programs theory. Evaluation tests the logic and results of program theory. While a theory may adequately describe what the project aimed to achieve, evaluation also investigates negative impact. Evaluation considers the possibility that at least some results may have occurred regardless of program presence. In evaluating the coherence, consistency and evaluability of a program theory, the assumptions—diagnostic, prognostic, and transformational—can play a key role in framing evaluation questions. It is for this reason that many evaluations conducted for community-based projects are better for hypothesis generation about services than for hypothesis testing (Leviton 1994), since they may involve a lot of unearthing of tacit program assumptions and refining of programs theory to make a plausible connection between the program and measured outcomes.

Assumptions, whether explicit or not, are the glue that hold a programs theory together. If the assumptions are fuzzy or faulty, it means that the program's theory is weak. Thus, any evaluative assessment of a program's theory should have at its core an assessment of a program's assumptions. In her paper "Which Links in Which Theories Shall We Evaluate?" Weiss (2000, pp. 35–45) illustrates how the transformational assumptions (the causal links between one program result and another) are the "theory" of a program. She uses an example of a theory for a job training program depicting successive stages of participants' behavior. This theory includes the stages of implementation as well as outcomes at various levels of change in/among program participants. The theory outlines a range of program milestones: for example, advertising of the training program; enrollment of eligible youth; provision of occupational training to the youth (accessible location, matches labor market needs and carried successfully); youth internalizing values of regular employment and behavior on the jobs; referral of youth to suitable jobs; to helping youths to make transition to work and stay in their jobs (p. 36).

While the program milestones—both the outputs and outcomes—might be and are usually a focus of the evaluator's attention, the links between those milestones are essential to examine the program's theory. For example, "whether trainees learn the skills taught, whether learning the skills leads to the search for a job, whether the search for a job leads to interviews with prospective employers, whether interviews lead to getting hired…." (p. 36). Such linkages, according to Weiss, are "assumptions about how the program is expected to succeed" (p. 36). Beside the more obvious linkages between outcomes, there are several other assumptions on how stakeholders expect the program to succeed: "perhaps the youth are rational enough to want to acquire skills that will help them get ahead in the job market, or perhaps program staff instill a sense of group esprit and a sense of excitement about the benefits of work that support the youth in staying in the

program" (Weiss 2000, 36). These assumptions are what Weiss calls "the real program theory" which when put together with the program activities-implementation theory; make the 'theory of change' (Weiss 2000, 37)

Weiss (2000) argues that a theory-based evaluator has to dig to uncover the implicit assumptions since there are always multiple views about what will make the program successful. Remember, then, that the discussions that dig into 'buried' or implicit assumptions about what will make a program successful do not have to wait for evaluation. If they are made explicit right from assessment and program design, implementation might still generate understanding around these assumptions, and evaluations may not meander around into digging these assumptions from the stakeholders, an exercise that can significantly eat into evaluation time and resources. (It suffices to mention that even when such assumptions are made explicit earlier on in the program, it is worth checking if they have change and if new assumptions have emerged.)

Janssens and Wolf's (2010) evaluation of education governance in the Netherlands is a good example of ex-ante evaluation of policy programs. They demonstrate how a specific policy might not achieve its objectives and illustrate the elements of the policy, which need improvement. They reconstruct and evaluate the theory behind the policy program that aims to increase productivity of schools in the Netherlands, the policy of "educational governance." Based on a combination of measures of internal quality assurances as well as accountability and supervision, the evaluation clearly shows a flaw in the theory underlying the policy and program, which threatens its potential effectiveness. Furthermore, there is a risk of contrary and incompatible interests among actors, as well as some practical reasons why the program might not work. Adjustment of the program is recommended before implementation

They identify five assumptions that underpin the program and review the validity/plausibility and empirical tenability of all the five assumptions by comparing them with research literature. The Table (11.1) outlines the results of their evaluation.

In their evaluation, a number of assumptions are reviewed, and when they are not supported by literature, they are considered to be weak and there was a resultant adjustment of program's theory before full-scale implementation.

Types of Evaluations that Test Assumptions of a Program Theory

Intervening mechanism, causative and generalization evaluations are some of the evaluations that focus on testing program assumptions. Chen (2005) refers to as one that intervening mechanism evaluation examines the underlying processes mediating between treatment and outcomes. Causative evaluations assess the causal processes underlying a program.

Table 11.1 Results of an evaluation of five assumptions of a program theory

Assumption	Evaluation result
1. Systems for internal quality assurance enable schools to generate an over all picture of their quality and to share it with their internal supervisors and their constituents	Assumption is currently implausible, because (a) quality assurance and self-evaluation are still in their infancy, and (b) accountability documents are deficient, partly because of the conflicting aims of self-evaluation (i.e. Improvement versus accountability).
2. When schools are accountable to their directly involved parties (horizontal accountability), these practices are able to exercise a positive influence on the self-regulation of the schools	(a) Accountability is important, as it encourages schools to improve their quality, and (b) the availability of accountability information should not be expected to increase the influence that people can exercise on the way schools are run.
3. School quality is improved by both external and internal supervision	External supervision works to some extent; internal supervision might work even better, as it is less sensitive to strategic behavior.
4. Internal supervision should be conducted professionally, independently, and in consideration of governmental goals.	The professionalism and independent nature of internal supervision should be seriously doubted, because in most case internal supervisors are lay persons appointed by the school board. The extent to which the governmental interest is considered by internal supervisors is low.
If schools generate accountability according to their internal quality assurance systems, t inspector will be able to coordinate their supervisory activities accordingly.	research findings question this because it is challenging to coordinate school accountability with external supervision because of differences in goals and the need for information (public accountability versus school improvement).

Generalization evaluation concerns the relevance of evaluation results to future application. It is difficult to generalize evaluation findings to settings where assumptions are different from those in which the evaluation has been done. Thus, generalization evaluation may have at its core a comparison of assumptions in the implementation domain to those made in the generalization domain—where the conclusions will be generalized: e.g. an evaluation that proves that school meals increase student's attendance may be difficult to generalize in communities with divergent livelihoods—e.g. one urban and another rural with families (and children) all involved in cultivation.

A number of techniques have been used to test both causal and intervening mechanism hypotheses. Scriven's General Elimination Methodology (GEM) identifies other possible causes of the observed change in outcomes, and discusses whether they could be responsible—see for example Scriven (2008). Other methods include Path Analysis (structural equation modeling), and pattern matching.

Path Analysis

Path analysis is one of several causal modeling procedures that might be used to represent and test program theory in evaluation (See Schumacker and Lomax 1996; Mitchell 1993; Maruyama 1998; Kelloway 1998).

The method is characterized by the analysis of non-experimental data in the absence of experimental controls by using variables implied in relationships with observable variables. It involves use of alternative causal models that imply patterns or relationships among variables. The underlying causal reasoning is made explicit in the form of structural equations.

The next step is developing a path diagram that displays the order in which variables are assumed to affect one another, and then regression models are generated to test the relationships among program variables.

Pattern Matching

Pattern matching is used to mean the correspondence between a theoretical or conceptual expectation pattern and an observed or measured pattern in the data (Marquart 1990) also see McLinden and Trochim (1998); Trochim (1987), (1989); Trochim and Cook (1992) etc. The approach is based on the interaction between the program's theory and observed program effects. Depending on the nature of the program, the program participants, and what is actually measured, this approach generates patterns in the program theory and outcome variables, and ascertains the degree of correspondence between or among these patterns.

In an employer-sponsored child care program, Marquart's (1990) evaluation focused on employees' attitudes and behaviors that affected their performance and value to the organization, and not on the effects of the childcare program on the children or parent–child relationship. The evaluator tried to obtain health administrators' implicit theories of this program. Concept mapping was used to understand administrators' perceptions of the relationship between the program and pertinent concepts, and causal ratings were used to obtain administrators' perceptions of the causal relationship between the program and its intended outcomes. Causal mapping was used to identify other organizational programs that were possible causes for the outcomes. The outcome pattern match assessed the correspondence between the order of effects predicted by health administrators and the actual order of effects found in the program evaluation. The health administrators predicted satisfaction with childcare and job satisfaction would be most affected and the absenteeism and turn over would be least affected by the program. The organization's personnel policies and supervisors' personnel practices were not perceived to be program effects. The evaluation data indicated that strong program effects (with statistically significant differences between participants and non-participants) were on recruitment, stress about childcare, and satisfaction with

childcare arrangement. Weak program effects (without significant group differences) were found on other variables. There was limited correspondence between administrators' predicted order of effects and the actual order of effects in the program evaluation. The implied theory of causal effects of the program was not supported by empirical data.

In this program evaluation, Marquart (1990) illustrates a focus, not on how all program milestones (such as effects on child wellbeing) were being achieved, but rather, how the stakeholders' "implicit" expectations around causal links correlated with the actual measured casual links. As Marquart's study found out, there was a disconnect between what stakeholders expected and the reality. Although stakeholders' expectations or implicit assumptions were not written down as coded rules, they would definitely influence the decisions made in managing the program, and unearthing as well as testing these assumptions through the pattern matching approach was used to later enrich the program theory and discard "faulty" implicit assumptions. Hence bring greater convergence between stakeholder and scientific theory.

Assumptions in Evaluating the Counter-Theory

A counter-theory according to Davies (2010) is an alternative theory of what has happened with the existing intervention. This is not the same as the counterfactual, which means thinking about something that did not happen. Exploring the counter-theory also involves ruling out plausible rival explanations, which is also a way of identifying and addressing program theory failure (Stame 2010). Since program theory represent "intended" results chain, it can be argued that the counter-theory may represent the "unintended" consequences.

Very often program evaluators talk about ''unintended effects." Chen and Rossi (1980, p. 107) argue that "every program has some effects". What do we mean by unintended effects? Merriam-Webster's dictionary defines 'intention' as a determination to act in a certain way.

Unintended effects have sometimes been defined as the 'unwanted' effects (e.g. University of Oxford 2010), but they may not necessarily be 'unwanted'. Unintended effects capture the idea that program stakeholders may not and cannot foresee everything that can happen in and after executing a program. There is uncertainness and lack of control, a very common attribute of complexity. This very attribute of complexity necessitate that programs are designed with elaborate theory. The presence of theory does not guarantee that the unexpected will not occur. In fact, theory may systematically position stakeholders for big surprises, especially if it is based on invalid assumptions. The role of theory is to reduce the level of surprises and to prepare stakeholders to systematically learn from the 'surprises'. In some cases, this entails that there will be no surprises, if the theory provides for multiple factors that could be at play to influence the success and failure of a program (unintended effects are not unexpected).

Evaluators need to be concerned about unintended consequences because they limit our ability to detect and explain the outcomes of the programs we evaluate (Morell 2005), and to know how to deal with unintended consequences, we need to know where they come from.

Morell (2005) categorizes unintended consequences into three: (a) the unforeseen and (b) the unforeseeable and (c) overlooked consequences. Unforeseen consequences exhibit less uncertainty than the unforeseeable consequences. "Overlooked consequences" may be known but deliberately ignored for practical, political, or ideological reasons. Unforeseen consequences emerge from weak application of analytical frameworks and from failure to capture the experience of past research. Reasons for unforeseeable consequences stem from the uncertainties of changing environments combined with competition among programs occupying the same "ecological niche". Tactics for dealing with the unforeseen include appreciation of life cycle behavior; the use of interdisciplinary teams; search for insight from past research; planning methods, such as alternate scenario planning, assumption-based planning, and back-casting; limiting temporal and causal distance between intervention and outcome; and group process techniques for managing high-conflict (Morell 2005).

One way of studying unintended effects of programs is to develop evaluation tools that widen observations beyond expected outcomes (usually through flexible qualitative approaches). Another way is to develop program theory that focuses on a wide range of possibilities (or the absence of those possibilities) such that the unintended effects are not unexpected. One of the ways to ensure that program theory captures 'unintended effects or relationships' is found in Chen.

Chen (2005) discusses the intervening mechanism evaluation approach, which assesses whether the causal assumptions underlying the program are functioning as they have been projected by stakeholders (p. 241). Its basic task is to assess the 'change model' aspect of the program theory. It includes determinants intervening between intervention and outcomes. One form of the intervening mechanism approach is the dynamic model. The dynamic model, which assumes that multi-directional, reciprocal causal relationships exist between the determinants, and the outcome, is reciprocal rather than one way. Other forms of intervening mechanism evaluation include: the multi determinant model with sequential order—two or more determinants are aligned in a causal order; multi determinant model with no sequential order—two or more determinants, each affected by the intervention, or affecting the outcome, but in no particular sequence), a one determinant model—involving a single determinant.

According to Chen, this approach of evaluation (intervening mechanism evaluation approach) is used to (a) assess whether causal chains functioned as expected, and how this might have contributed to the success or failure of the program; (b) when action theory fails, meaning that an intervention did not affect its determinant, yet the conceptual theory is sound, (c) when conceptual theory fails, although the intervention activates its determinants, the determinant may fail to affect the outcome, or (d) when both the action and conceptual theories fail.

The Integrative Process Outcome Approach

This approach involves a systematic assessment of both the crucial assumptions beneath implementation and the causal processes of the program because such a thorough analysis of potential pathways enlightens stakeholders as to how their system truly operates" (Chen 2005, p. 241).

These two are approaches of the counter-theory, looking at dynamics differently from the program theory. These show that programs do not necessarily work according to the theory we design. Therefore, stakeholders need to think through alternative explanations of how the program has worked in practice.

The examination of assumptions (including diagnostic, prognostic, and transformational assumptions) can be a crucial part of the integrative approach of examining unintended effects'. This begins with stating and reviewing both the rationale and risks to assumptions, and continues with monitoring and evaluating the validity of diagnostic, prognostic, and transformational assumptions along with regular process and outcome monitoring.

Since program effects are conceived in relation to problem causes, diagnostic assumptions can be reviewed to determine whether the originally conceived problem causes were valid, or alternative causes can be reviewed.

Prognostic assumptions can be reviewed to determine if there are actions or interventions that would have been more appropriate to the problem and whether elements of such interventions have been implemented or not.

Transformational assumptions can be reviewed to make sure that the conceived linkages between outputs, short-term outcomes and intermediate outcomes, and long-term outcomes are valid, and if alternative linkages can be conceived.

References

Chen, H. T. (2005). *Practical program evaluation: assessing and improving planning, implementation, and effectiveness*. Newbury Park: Sage publications.

Chen, H. T. (2006). A Theory-driven evaluation perspective on mixed methods research. *Research In the Schools (Mid-South Educational Research Association), 13*(1), 75–83.

Chen, H. T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces, 59*(1), 106–122.

Connell, J. P. & Kubisch, A.C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives: Progress, prospects, and problems. In K. Fulbright-Anderson, A. C. Kubrisch & J. P. Connell (Eds.), *New approaches to evaluating community initiatives ,theory, measurement and analysis* (vol. 2). Washington: Aspen Institute.

Davies, R. (2010). *Counter-factual and counter-theories*. Retrieved 01/29/2012 http://mandenews.blogspot.com/2010/10/counter-factuals-and-counter-theories.html.

Donaldson, S. I., & Gooler, L. E. (2002). Theory-driven evaluation of the work and health initiative: A focus on winning new jobs. *American Journal of Evaluation, 23*(3), 341–347.

Funnel, S. C. (2000). Developing and using a program theoryprogram theory matrix for program evaluation and performance monitoring. *New Directions for Evaluation, 87*(Fall), 91–101.

Green, B. L. & McAllister, C. (1998) Theory-based, participatory evaluation: A powerful tool for evaluating family support programs. In *The bulletin of the national center for zero to three* (pp. 30–36) Feb/March 1998.

Janssens, F. J. G., & De Wolf, I. F. (2010). Analyzing the assumptions of a policy program: An ex-ante evaluation of ''educational governance'' in the Netherlands. *American Journal of Evaluation, 30*(3), 330–348.

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling*. Thousand Oaks: Sage.

Leviton, L. C. (1994). Program theory and evaluation theory in community-based programs. *American Journal of Evaluation, 15*(1), 89–92.

Marquart, J. M. (1990). A pattern-matching approach to link program theoryprogram theory and evaluation data. *New Directions for Program Evaluation, 47*, 93–107.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks: Sage.

Mayne, J. (2011) Contribution analysis: Addressing cause effect. In K. Forss, M. Marra & R. Schwartz (Eds.), *Evaluating the complex: Attribution, contribution, and beyond* (pp 53–96). New Brunswick: Transactional Publishers.

McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your performance story. *Evaluation and Program Planning, Elsevier Science, 22*(1), 65–72.

McLinden, D. J., & Trochim, W. M. K. (1998). Getting to parallel: Assessing the return on expectations of training. *Performance Improvement, 37*(1998), 21–26.

Mitchell, R.J. (1993). Path analysis: pollination. In S. M. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (pp. 211–231).New York: Chapman and Hall.

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

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Schalock, R. L., & Bonham, G. S. (2003). Measuring outcomes and managing for results. *Evaluation and Program Planning, 26*(3), 229–235.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah: Lawrence Erlbaum Associates.

Scriven, M. (2008). A summative evaluation of RCT methodology: & an alternative approach to causal research. *Journal of Multi Disciplinary Evaluation, 5*(9), 15–24.

Shaw, I., & Crompton, I. A. (2003). Theory like mist on spectacles, obscures vision. *Evaluation, 9*(2), 192–204.

Stame, N. (2010). What doesn't work? three failures, many answers. *Evaluation, 16*(4), 371–387.

Tilley, N. (2004). Applying theory-driven evaluation to the british crime reduction program: The theories of the program and of its evaluations. *Criminology and Criminal Justice, 4*(3), 255–276.

Trochim, W. (1989). Outcome pattern matching and program theoryprogram theory. *Evaluation and Program Planning, 12*(4), 355–366.

Trochim, W., & Cook, J. (1992). Pattern matching in theory-driven evaluation: a field example from psychiatric rehabilitation. In H. Chen & P. H. Rossi (Eds.), *Using theory to improve program and policy evaluations* (pp. 49–69). New York: Greenwood Press.

University of Oxford (2010). *Inventory of measures, typology of non-intentional effects and a framework for policy packaging.* Retrieved 11/13/11 http://optic.toi.no/mmarchive_getfile.php?mmfileid=14934&CPMMFILEID_URL_WYSIWYG_TOKEN=1.

Weiss, C. H. (1993). Where politics and evaluation research meet. *American Journal of Evaluation, 14*(1), 93–106.

Weiss, C. (1997a). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation, 76*, 41–55.

Weiss, C. (1997b). How can theory-based evaluation make greater headway? *Evaluation Review, 21*(4), 501–524.

Weiss, C. (2000). Which links in which theories shall we evaluate? *New Directions for Evaluation, 87*, 35–45.

Chapter 12
Challenges of Explicating Assumptions

Abstract The process of *explicating* assumptions requires an analytical stance
that is different from the empathetic, responsive, and intuitive stance of many
stakeholders (Weiss New Approaches to Evaluating CommunityInitiatives, The-
ory, Measurement and Analysis, vol 2, pp. 65–92, 1995). Because developing and
questioning the underlying assumptions of a program's theory can be threatening
to participants, it is important that the process be done in a sensitive and collab-
orative way, so that it strengthens the program without dampening enthusiasm or
weakening stakeholder commitment. Explicating implicit assumptions is some-
thing stakeholders may be uncomfortable with for different reasons. Some of these
are outlined below.

The Nature of Implicit Assumptions

The biggest challenge with explicating implicit assumptions is that they are
implicit: as discussed in Part II of this book (implicit) assumptions are embedded
in paradigms which guide the way we view the world around us. They are formed
in very subjective ways as we interact with people around us and the physical,
socio-economic, and political environments in which we live. They are cemented
into belief systems. We may not be aware of them or even when we are aware of
them, we may not know how and when they influence how we look at the world
around us. When we do, we tend to defend, justify, and rationalize them as truth,
with a host of explanations.

The fact that stakeholders may not be aware of their implicit assumptions does
not mean they should not explicate them. It is necessary since these assumptions
influence program choices. It is important that a facilitator recognizes this

difficulty and designs adequate reflective tools that enable stakeholders to recognize, and verbalize these assumptions.

Explicating Implicit Assumptions May Not Be Desirable

Some communities may have a culture of being implicit, and being direct may be viewed as a form of rudeness. This may be true with a Rukiga proverb (from western Uganda), "if something is of value, such as truth, you should carefully dress it and present it in an envelope.." It may require strong facilitation skills to get stakeholders to explain what they believe are or are not the causes of a problem. The other problem is that people sometimes fear to be so specific, as they are not sure what they are saying is accurate. They may fear to be held could responsible for failure—"this is what you said and it is not turning out to be true".

Explicating Implicit Assumptions is an Abstract, Intangible and Not Concrete

Weiss (1995) highlights this problem and argues that many program people will find the process uncongenial. Doing something is less demanding and less strenuous in terms of thinking than discussing in an explicit and critical way what it is one is doing (Scriven 1986). The process of explicating assumptions requires an analytical stance that is different from the empathetic, responsive, and intuitive stance of many development practitioners. Stakeholders may find it difficult to trace the mini-assumptions that underlie their practice, be uncomfortable with the attempt to pull apart ideas rather than deal with them in broad themes, and may feel that the process is not adding value and question the utility of the approach.

The Implications of Explicating Assumptions

According to Kaplana and Gharret (2005), stakeholders may resent being asked to question long-held beliefs, or to provide evidence to support their work. Moreover, the questioning of program assumptions may lead to the need to reallocate resources and responsibilities that have already been allotted. Such a process potentially raises issues not only within the community but also with funders. This exercise of reexamining assumptions usually points out gaps in the program's theory and may imply that it should be revised drastically. This is a good thing, but may be difficult after funding arrangements have been agreed upon (Kaplan and Garrett 2005).

Because developing and questioning the underlying assumptions of the program model can be threatening to participants, it is important that the process be done in a sensitive and collaborative way, so that it strengthens the program without dampening enthusiasm or diminishing gut level commitment (Kaplan and Garrett 2005).

It is important to emphasize that the value in spelling out these assumptions is to clarify stakeholders' understanding of the program that they will be designing, and make it easy to learn from the program once it is implemented. The assumptions represent stakeholders' current understanding, which does not have to be rigid or fixed. These assumptions can be validated or challenged along the way. New assumptions will emerge and will guide the program and they will be validated or challenged within the context of the program delivery. This is a form of intentional learning.

The Limitations of Explicating Implicit Assumptions

Its Utility Depends on Stakeholders' Ability to Design and Implement Programs Appropriately, and The Evaluator's Ability to Formulate Appropriate Questions

Explicating implicit assumptions is not all there is in development programs. It is just as small but a critical portion of it. Its utility depends on the ability of program stakeholders to question the validity of the explicated assumptions and revise a program design accordingly or where design revision is not possible, to interpret monitoring and evaluation information within the context of these assumptions. A program evaluator needs to be able to formulate and measure evaluation questions with respect to these assumptions and interpret information as such.

Not every assumption can be validated and not every faulty assumption has to be discarded. The importance of explicating assumptions is not limited to validating and discarding. It is meant to increase understanding; to guide decisions with greater accuracy and hence greater development effectiveness. Also, making assumptions is not necessarily a result of sloppy thinking. The danger is in the lack of awareness that one is making an assumption—on the part of the one who is assuming, as well as others most affected by the assumption.

Unknown Unknowns

Assumptions function within a program's theory. This theory describes, prescribes, and predicts how programs will perform. But what if the stakeholders do not get it right? In their paper "Theory, Like Mist on Spectacles, Obscures Vision" Shaw and Crompton (2003) provide a caution that even after a program's

theory and underpinning assumptions have been elaborately stated, it is important to remember that in choosing a theory we also exclude other theories and exclude areas of information that lie outside of the particular theoretical framework of choice. This implies that the formulation and testing of program theory and assumptions is a practice that continues through implementation and evaluation. To some practitioners, such a detailed account may be impractical. But Weiss (1972) in McCoy and Hargie (2001, p. 317) has argued that "one program's over-elaboration is another program's clarification". Attributing effects or changes in program participants to the program can only be done by understanding the process that is meant to bring about those changes.

Prioritizing Assumptions or Theories to Focus on

At the same time, the evaluator and stakeholders could get lost in a maze of assumptions and program outcomes that have to be evaluated, and this could also make the evaluation expensive. It is important to prioritize. Weiss (2000) gives some criteria to follow:

(1) The first criterion is the beliefs of program stakeholders—What do these groups assume are the pathways to desired program outcomes; the causal linkages among mini-steps to these outcomes; and what stakeholders believe is crucial to eventually bring about the outcomes.
(2) Is the theory plausible? Can the program design be implemented with fidelity? Will the program beneficiaries respond according to the program's expectations?
(3) Where is the lack of knowledge on the program? If it is assumed that giving information will increase knowledge and increased knowledge would lead to behavior change, it may be important to determine on which links in this chain of results is information available, and where information is not available.
(4) What is the centrality of the theory to the program? The different sub theories and assumptions may operate together or separately within a program but they may not all hold equally important roles. Some of them may be more pivotal to the program's success than others. Such theories and assumptions need to be prioritized. As Weiss suggests, "it seems wise to spend resources to study a particular on which the program most basically rests. If the program is predicted on the assumption that what keeps youth enrolled in the full training program is the support of their peers, then the assumption warrants investigation" (Weiss 2000, p. 44). Even though some theories, linkages, or assumptions may be more pivotal to a program's success than others, stakeholders may not have the same degree or uncertainty on all of them. Again, it is important to prioritize studying assumptions that are not likely to be empirically supported, or on which prior studies have produced no, weak, or conflicting evidence.

The Bottom Line

Explication of assumptions is so vital even at evaluation. The suggestions on how to prioritize assumptions for study outlined above apply mostly at evaluation. But even during program design, it is necessary to define boundaries of what is essential for the success of a program.

In 2011, I was facilitating an assumptions explication exercise in western Sri Lanka when one participant asked me: "what if after you explicate all the crucial assumptions; you realize that they are faulty and your program strategies are therefore; do you cancel the program?" You might have already wondered about this as you read various parts of this book. In fact, what to do with the program after the assumptions are explicated depends on:

(a) The stage in the program cycle at which assumptions are explicated. The earlier, the more flexible and the easier for the stakeholders to make changes in program plans and strategies.
(b) The scope of stakeholder involvement: If the stakeholders who realize the need to revise program strategies and hence the assumptions on which strategies are premised are different from stakeholders who make key program decisions, revising the program may be difficult.

Thus, assumptions explication should begin early in the program design process; continue through program implementation and monitoring; and be a part of the program's evaluation and learning. There may be different ways of explicating assumptions; different types of assumptions; different ways of dealing with them; in fact different programming contexts may require a different selection of tools and facilitation techniques. The bottom line is, stakeholders need to explicate their assumptions of the root causes of problems and needs; the most appropriate strategies to address these; and the most appropriate program strategies to address these problems; and their expectations of how program interventions will eventually alleviate the problem by producing desired outcomes. Consider again the story of the young girl and the pregnant woman. The two brought different assumptions to the argument on the bulging stomach, which caused a conflict between them. Even if the two assumptions were different (and the girl's assumptions were faulty for that matter); that probably would not have caused this conflict, if the girl accompanied her statement: "I know what you did" with her assumptions. The explication of assumptions would make evaluations easier to do; and thus easier to learn from.

References

Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning, 28*(2005), 167–172.

McCoy, M., & Hargie, O. D. W. (2001). Evaluating evaluation: implications for assessing quality. *International Journal of Health Care Quality Assurance, 14*(7), 317–327.

Shaw, I., & Crompton, I. A. (2003). Theory like mist on spectacles obscures vision. *Evaluation, 9*(2), 192–204.

Scriven, M. (1986). New frontiers of evaluation. *American Journal of Evaluation, 7*(7), 7–39.

Weiss, C. H. (1972). *Evaluation research—methods for assessing program effectiveness*. Englewood Cliffs: Prentice Hall.

Weiss, C. H. (1995). Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In K. Fulbright-Anderson, A. C. Kubrisch & J. P. Connell (Eds.), *New Approaches to Evaluating Community Initiatives, Theory, Measurement and Analysis* (pp. 65–92, vol. 2). Washington: Aspen Institute.

Weiss, C. (2000). Which links in which theories shall we evaluate? *New Directions for Evaluation, 87*, 35–45. Fall.

Index