A META-EVALUATION OF THE SUCCESS CASE METHOD APPLIED TO A LEADERSHIP DEVELOPMENT PROGRAM

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY OF RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

KEVIN ROBERT ENGHOLOM
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY MAY 2016

APPROVED: ________________________________
Cary Cherniss Ph. D.

______________________________
Bradford Lerman Psy. D.

DEAN: ________________________________
Stanley B. Messer Ph. D.
Copyright 2016 by Kevin Robert Engholm
The study explores meta-evaluation as an approach that corporate learning functions can employ to assess the efficacy of a given evaluation method. To that end, an internal meta-evaluation was conducted to determine the utility, feasibility, propriety and accuracy of an already completed Success Case evaluation of a leadership development program within a global bank. Twenty-one subjects from the company’s Human Resources department, including the researcher, participated in the meta-evaluation. The researcher personally recruited the subjects based on their involvement with the leadership development program’s design and deployment. Data were collected via online questionnaire, semi-structured interviews, and a review of archival data. The meta-evaluation findings suggest that the Success Case evaluation met the overall standard of propriety to a “very great extent,” and the standards of accuracy, feasibility and utility to a “great extent.” Specifically, while the participants in the study agreed with the Success Case evaluation’s primary conclusion that there were opportunities for the program to have greater business impact, they also identified limitations in the evaluation’s recommendations to improve the program and increase manager engagement. In addition to the efficacy of the Success Case Method using meta-evaluation criteria, the study discusses the opportunities and limitations of meta-evaluation as a potential approach to enable organizations to develop more robust, effective and comprehensive evaluation strategies.

*Keywords: Success Case Method (SCM), Brinkerhoff, Meta-Evaluation, Meta-Evaluation Standards, Learning, Kirkpatrick, Training Transfer, Impact, Accuracy, Propriety, Utility, Feasibility*
Dedication

This dissertation is dedicated to the memory of my father, Robert Engholm. Our home was filled with many books, lively conversation, and encouragement to “redeem the time.” From Dad I learned the value of discipline, hard work, and commitment. My Mom, Aloha Engholm, has also been a constant source of inspiration; her curiosity and passion for learning has only intensified over the years. Together, they created a home environment of unconditional love that cultivated in me the desire and belief that perseverance pays off. Although I was unable to complete this dissertation before Dad passed away in 2014, I’m confident that knowing that I had finished would have made him proud.
Acknowledgements

This dissertation would not have been possible without the ongoing support and forbearance of Dr. Cary Cherniss. Cary’s efforts to be available, review drafts, and provide input went well beyond the call of duty. I’ll always be grateful to have had the honor and privilege of working with such an accomplished scholar, teacher, and human being.

Similarly, I’m grateful to Dr. Brad Lerman, who served as a mentor to me as a first-year student at GSAPP and has continued to be a role-model and source of thoughtful advice and support. I’d also like to thank Dr. Charlie Maher who introduced me to the world of Program Planning and Evaluation.

There are many friends and colleagues who have cheered me on, even when I’m sure they privately doubted I would finish: Eric Berger, Andy Burt, Ruth and Steve Carlsen, Emily Dancyger King, Ariel Frank, Ted Freeman, Terry Hogan, Tzivya Katz, Steffen Landauer, Cristi Lockett, Chandra Mallampalli, Chris and Claudia Matthews, Pat Mandica, Thor Mann, Paul McKinnon, Keith Miles, Amanda Rose, Melanie Stopeck, Fred Stern, and all my colleagues from LEAP and ICG/GF Learning. My three sisters Karen, Kathy, and Kay (and respective spouses) and have never wavered in their love and unconditional support.

No one, however, deserves more gratitude than my beloved wife Elvira and our three precious daughters: Miranda, Rebecca and Gabriela. They have grown up with this project as backdrop to their lives, as it has been a constant companion on vacations and weekends since we all can remember. Lastly, I want to acknowledge my Lord Jesus Christ, whose grace enabled me to complete this work (2 Cor. 12:9).
# TABLE OF CONTENTS

**ABSTRACT** .......................................................................................................................... ii

**DEDICATION** ..................................................................................................................... iii

**ACKNOWLEDGEMENTS** ..................................................................................................... iv

**LIST OF TABLES** ............................................................................................................... ix

**LIST OF FIGURES** .............................................................................................................. x

**CHAPTER I: Introduction** .................................................................................................... 1

  - Development of evaluation in public education and social programs ...................... 3
  - Development of evaluation in business and industry ................................................ 5
  - Practitioner-related barriers to effective evaluation .................................................... 7
  - Leadership training represents a unique evaluation challenge ................................ 10
  - Transfer of training ........................................................................................................ 10
  - Brinkerhoff’s Success Case Method as a promising approach for evaluating leadership development programs ........................................................................................................ 12
  - Meta-evaluation as a tool to bridge the gap between research and practice .......... 13
  - Which standards should guide a meta-evaluation? ...................................................... 15
  - The study .......................................................................................................................... 19
  - Research Questions ........................................................................................................ 20

**CHAPTER II: Background of the organization, leadership program, and the Success Case Method (SCM)** ........................................................................................................ 22

  - The bank ........................................................................................................................... 22
  - Genesis of the leadership program ................................................................................ 23
  - Description of the training program .............................................................................. 24
  - Pre-work ............................................................................................................................ 27
DISSERTATION: APPLIED META-EVALUATION

The program.................................................................................................................. 29
The program’s deployment and evaluation................................................................. 30
The Success Case Method............................................................................................ 31
Summary of the Success Case Evaluation study of the leadership program .......... 38
Success Case Evaluation conclusions and recommendations ............................... 39

CHAPTER III: Method.................................................................................................. 42
The original study (Success Case Evaluation) ......................................................... 42
The present study (Meta-evaluation) ........................................................................ 42
Meta-evaluation participants ...................................................................................... 42
Informed consent and confidentiality.......................................................................... 43
Procedure .................................................................................................................... 44
Analysis ....................................................................................................................... 50

CHAPTER IV: Results .................................................................................................. 53
Electronic survey ........................................................................................................ 53
Semi-structured interviews ....................................................................................... 55
Results in relation to research questions .................................................................... 56

CHAPTER V: Discussion.............................................................................................. 68
Summary of findings .................................................................................................. 69
Burke-Litwin Model of organizational change ......................................................... 77
Overall value of the meta-evaluation for the organization ....................................... 81
Changes in participants’ views over time ................................................................. 87
Changes in participants views and organizational maturity .................................... 91
Role of the investigator............................................................................................... 92
Limitations of the study ............................................................................................ 94
DISSERTATION: APPLIED META-EVALUATION

Contributions of the study and implications to the field........................................95

Implications for future research..................................................................................96

References.....................................................................................................................100

APPENDIX A: Informed Consent Form ......................................................................106

APPENDIX B: Online Questionnaire ..........................................................................108

APPENDIX C: Individual Interview Protocol..............................................................119

APPENDIX D: Learning Impact Map of the leadership program .................................124

APPENDIX E: List of documents for archival review ................................................126

APPENDIX F: Guiding Questions for Archival Review ..............................................127

APPENDIX G: Success Case Impact Survey ................................................................129
LIST OF TABLES

TABLE 1: Meta-Evaluation Quantitative Survey Results........................................54
LIST OF FIGURES

FIGURE 1: Burke-Litwin Model of Organizational Performance and Change ..................80
Chapter I: Introduction

Over the past quarter century, Ray Stata’s (1988) statement that “the rate at which individuals and organizations learn may become the only sustainable competitive advantage” (p. 64) has served as a rallying cry for corporate training functions. The assertion simultaneously captures the learning profession’s highest aspirations while serving as a painful reminder that this vision is still a distant reality.

Until recently, the training function has resided in the margins of most companies, viewed as providing a tertiary benefit or support to employees, but not acting as critical to the fulfillment of the organization’s strategy. Gradually however training has begun to secure a “seat at the table,” increasing both the visibility and expectations from the training function. As the ASTD’s 2004 State of the Industry Report noted (Sugrue and Rivera, 2005, p. 5):

The status of the learning organization has been elevated as more and more organizations appoint a chief-level officer with responsibility for learning who reports directly to the CEO rather than through HR; but with elevated status come elevated expectations. These expectations are translated into mandates to “run learning like a business,” “demonstrate the value of learning,” and “drive organizational performance.”

This heightened focus has continued despite the Great Recession of 2007-2009, with investment in employee learning in the U.S. alone reaching $164.2 billion in 2012 (Miller, 2013) and average direct expenditure per employee estimated at $1229 in 2014 (Ho, 2015).

Despite the increased optimism and investment, many leaders of corporate training functions, rather than having a “seat at the table,” still find themselves in the waiting room for two fundamental reasons. The first reason is a failure of the training to be translated into the desired performance and outcomes (ASTD, 2006; Baldwin and Ford, 1988; Broad and Newstrom, 1992;
DISSESSATION: APPLIED META-EVALUATION
Cherniss and Goleman, 1998; Learning and Development Roundtable, 2009). The second reason lies in a failure in metrics and evaluation. This study’s premise is that the challenge of effective learning transfer (reason one) cannot be adequately addressed without an increased understanding derived through metrics and evaluation (reason two). Those training functions, which are unable to establish a compelling business case for the impact of their efforts, will continue to be vulnerable to the vicissitudes of the marketplace and the subjective perceptions of senior sponsors regarding the value rendered. The old adage, “Training is the first thing to go,” is frequently a reality. In tough economic times, judgment on the value of the training’s impact is rendered with or without solid evidence.

The demand for greater evaluation capability and accountability is not new. It has been a consistent theme in training literature since Donald Kirkpatrick first issued his clarion call for better evaluation in his seminal essays in Training and Development Magazine in 1959. In the first article of that series, Kirkpatrick quoted Daniel Goodacre from BF Goodrich as having said, “Training directors might be well advised to take the initiative and evaluate their programs before the day of reckoning arrives” (as cited in Kirkpatrick and Kirkpatrick, 2010, p. 3). That day has come.

Kirkpatrick’s articles awakened the field to this critical need for greater evaluation. Training professionals consistently report that measuring the business impact and other outcomes of leadership and executive development programs is one of their highest priority areas of interest and concern. Yet, despite this increased focus and awareness, progress has been limited.

In a Learning and Development Roundtable (2009) Learning Effectiveness Survey, only 33% of the managers surveyed either agreed or strongly agreed that “Learning & Development (L&D) is central to improving the performance of current employees.” The study further found
that 56% of these managers believed that employee performance would not change if L&D were eliminated today. In a 2009 survey Chief Learning Officer magazine conducted among its Business Intelligence Board, only 35% of respondents indicated they were satisfied with their organization’s learning measurement (Anderson, 2009). Similarly, in a joint study between the American Society for Training and Development (ASTD) and the Institute for Corporate Productivity (i4cp), only 25.6% of respondents believed that they received a “solid bang for their buck” when it comes to learning metrics (Bingham, 2009, p. 7). While companies today seem to recognize a problem exists, the same study reports that only 5.5% of the overall training budget is allocated toward its evaluation (Bingham, 2009). In a study conducted with 96 CEOs, Philips and Phillips (2010) reported that these senior executives are looking for data that demonstrate impact on the business and return on investment (“ROI”). While 96% of survey respondents indicated that impact was a measure that should be tracked, only 8% of the CEOs in the survey said that they were actually tracking this measure.

Why is there such a disparity between expectations relative to evaluation and actual practice? There are many potential answers to this question, but part of the answer lies in how the evaluation field has evolved along two parallel but largely non-intersecting paths in public and corporate education.

Development of Evaluation in Public Education and Social Programs

Evaluation has existed informally for millennia, but did not develop as a formal profession or area for academic research until the 1960’s when President Lyndon Johnson launched the “War on Poverty” and related Great Society programs. In 1963, the eminent educational

---

1 The War on Poverty is the unofficial name for legislation first introduced by United States President Lyndon B. Johnson during his State of the Union address on January 8, 1964. Johnson proposed this legislation in response to a national poverty rate of around nineteen percent.
psychologist Lee Cronbach published a landmark article entitled, “Course Improvement through Evaluation,” which encouraged evaluation of programs while still in design, stating that "evaluation used to improve the course while it is still fluid contributes more to improvement of education than evaluation used to appraise a product already on the market." (as cited in Madaus, Scriven, and Stufflebeam, 2000, p.105). This new paradigm, along with increased government expenditure and funding for the Great Society programs, called for greater accountability. The tipping point came in 1965 with Senator Robert Kennedy’s push to delay the passing of the Elementary and Secondary Education Act (ESEA) until it contained a clause ensuring that there would be an evaluation plan and summary report. As a result, every subsequent federal grant for programs began to require a formal evaluation plan and evaluation. The problem however was that initial evaluation quality proved to be inconsistent and relatively few individuals possessed the requisite understanding of evaluation as an applied discipline to meet this new demand.

As a response to this demand for formal evaluation, the first professional journals in evaluation began to appear in the 1970’s. Universities started to offer courses and programs specifically oriented toward building evaluation capability (Hogan, 2007). In 1974, the Joint Committee on Standards for Educational Evaluation was formed with a mission “to develop and implement inclusive processes producing widely used evaluation standards that serve educational and social improvement.” (Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A., 2011, p. xviii). In 1981, the first of three “Program Evaluation Standards” was published; subsequent revisions occurred in both 1994 and 2011. Two U.S.-based professional evaluation
associations, the Evaluation Network and the Evaluation Research Society, were established in the 1970’s. The former was made up of school-based evaluators, the latter with government and university-based evaluators. These two societies merged in 1985 to form the American Evaluation Association (AEA). Today, in addition to having over 7,000 professional members globally, has published its own Guiding Principles for Evaluators (2004)\textsuperscript{3}, and has chartered 55 separate Topical Interest Groups (TIGs)\textsuperscript{4}. These TIGs focus on specific areas and population groups for evaluation research and practice. It is instructive to note for the purposes of this study that only one of the 55 is specifically related to training evaluation in business and industry: The Business, Leadership and Performance TIG.

**Development of Evaluation in Business and Industry**

Development of evaluation in business and industry has followed a separate path. In Kirkpatrick’s landmark essays (1959) mentioned previously, he developed a four-level framework, which remains the prevailing model for evaluation in the training industry today. To wit, Bassi, Gallagher, and Schroer (1996) reported that 96% of companies use some version of Kirkpatrick’s framework to evaluate training.

Kirkpatrick’s motive (1998) was to substantiate the term “evaluation” by explaining four specific levels of approach. Level 1 describes the participant’s *reaction* to the training. Level 2 attempts to measure the level of *learning* achieved. Level 3 is focused on determining the extent to which behavior has *changed* as a result of attending the training program. Level 4 identifies the ultimate *results* that can be attributed to the participant’s having attending the training. It is important to note, however, that positive Level 1 and Level 2 results are not necessarily equated with satisfactory learning outcomes (Alliger and Janak, 1989).

\textsuperscript{3} Source: American Evaluation Association website, retrieved from www.eval.org/p/cm/ld/fid=51
\textsuperscript{4} Source: American Evaluation Association website, retrieved from http://www.eval.org/p/cm/ld/fid=11
Despite its prevalence, Bates (2004), citing Alliger and Janak, has cataloged critiques leveled against the approach, most notably that it presents an oversimplified view of training effectiveness. Specifically, he considers it to be incomplete, ignoring what happens before and after the training, that it makes assumptions of causal linkages between levels (e.g., if the trainee likes the training, Level 1, she is likely to learn more, Level 2), and that it makes the assumption that each subsequent level provides information of greater value (Bates, 2004, pp. 342-343).

While Kirkpatrick’s work has often been viewed as a method for training evaluation, it can more accurately be described as a heuristic or taxonomy to help frame key evaluative elements. This (mis)interpretation of Kirkpatrick’s model as a method (vs. a heuristic) has been unfortunate, as it has kept many practitioners focused more on devising measures to fit the taxonomy rather than on creating practical methods which would facilitate a more comprehensive understanding of training impact.

Jack Phillips cites this absence of how-to information as the genesis for the publishing of the Handbook of Training Evaluation Measurement Methods in 1983, which became the first U.S. book focused specifically on training evaluation (Phillips and Phillips, 2007). Philips also expanded on Kirkpatrick’s model to add a Level 5, which he defined as the “measurement of ROI,” which would supplant results to become the highest level of evaluation. With each successive rung in the ladder, greater effort and resources are required to extract the measurement. Therefore, the extent to which each level is measured in practice decreases as the level increases. This is borne out in the i4cp/ASTD study (Bingham, 2009) cited earlier, which estimates the following percentage breakdowns:

- 91.6% of organizations measure Level 1 reaction;
- 80% measure Level 2 learning;
The study found that Level 5 ROI was perceived as less valuable than having Levels 3 and 4 evaluation data. This oddity was mainly related to issues of stakeholders questioning the credibility of the ROI data. Most simply, ROI is the ratio of net benefits divided by costs expressed as a percentage. In order to be a credible calculation, there must be agreement on all of the variables to be taken into account and how these variables will be operationalized for measurement. Even when there is agreement on the variables to be measured, the challenge is to ensure that these benefits can be rightly attributed to the program. Frequently, there is no clear understanding of the assumptions behind a given ROI calculation, and even within the same organization, it is not difficult to imagine there being variability in how they are applied. This is not to say that ROI calculations are not worthwhile, but that they must be used judiciously given the effort required.

Kirkpatrick and Phillips have each made important contributions by shifting the focus of evaluation from one a focus on activity and opinion (e.g., reporting how many people had completed the training, how well was the training received, etc.) toward one that is more based on business outcomes.

**Practitioner-Related Barriers to Effective Evaluation**

In addition to the fact that corporate training evaluation has developed independent of the volume of research in the educational and public sector, another factor contributing to the poor state of corporate training evaluation has to do with those professionals who enter the field. Swanson (2005) has argued that those who have entered the training profession do not possess
either the mindset or the skills oriented toward measurement and evaluation. Moller and Mallin (1996) conducted a study of instructional design practitioners and found that while nearly 90% conducted end of course evaluations and 71% of these evaluate learning, only 44% of the group utilized acceptable techniques for measuring achievement. Of these, only 20% were able to correctly identify specific methods for results evaluation.

Kraiger, McLinden and Casper (2004) reported that less than one percent of the trainers they surveyed regularly read professional, peer-reviewed journals. Additionally, Hutchins, Burke, and Berthelsen (2010) concluded that most training practitioners are largely unaware of extant professional research. Their survey results suggest that training professionals seek knowledge mostly through informal learning (e.g., job experiences, discussions with internal and external training professionals, books, searching the Web), but they prefer to learn about training transfer in discussions with external trainers and academics. With this as a backdrop, it is not surprising that practitioners find themselves confused by evaluation approaches that they view as too complex to be practical or sustainable within the corporate environment (Phillips and Phillips, 2010).

Inadequate needs analysis is also an important factor. The Learning and Development (2009) Roundtable (LDR), in its Learning Effectiveness Survey, reported that while nearly three-quarters of participants in L&D programs reported very high degrees of learner satisfaction (with program, instructor, course design, materials, etc.), that this had no correlation to application back on the job. Astonishingly, they found that “more than half of all L&D programs suffer from a lack of relevance to day-to-day work, minimal emphasis on application during solution design, and low levels of learner motivation and manager ability to apply what has been learned.” (p. 3). The study’s authors recommend that the training design needs to focus on
learning and doing, while focusing on a detailed workflow analysis of potential participants in order to better identify the opportunities and obstacles for application. Additionally, engaging multiple stakeholders early in the process to identify key gaps and align around key outcomes for the training is imperative.

Adding to this challenge is a lack of resources, expertise, and organizational culture to support improved evaluation efforts (Desimone, Werner & Harris, 2002). As Chris Moore, writing for Chief Learning Officer Magazine (2009) has put it, the learning function is “simply too busy dealing with the day-to-day tactics of learning administration and delivery.” Where measurement is happening in large organizations, it is often only at the most basic level of evaluation: attendance, reactions to the program, and the associated costs. While this basic level is better than no evaluation, these measures offer few answers relative to the efficacy of the training and how it can be improved. Spitzer (1999) noted the ambivalence that frequently characterizes the attitudes of training professionals toward evaluation. Spitzer reports that training professionals themselves do not possess the confidence that their programs add value or have positive impact on their organizations. Spitzer found that training professionals tend to view evaluation more as a threatening referendum on their own effectiveness rather than a reflection of the whole system’s quality of performance.

Given the above, it is not surprising that evaluation is frequently considered as an afterthought as opposed to being an integral component of the program design process. It is illustrative that the best-known and most widely used program design model, known as ADDIE, originally was developed during World War II. ADDIE has held a place in the corporate training world since, largely due to its simplicity (Wang and Wilcox, 2006). The acronym ADDIE stands for Analyze, Design, Develop, Implement and Evaluate. The fact that evaluation is included as
part of the design process is a positive, but is still likely to be incorporated at the very end of the
cycle rather than being a central consideration from the outset. Given that this has been the
prevailing design model for training within business and industry, it is little wonder that
evaluation has often been relegated to little more than an after-thought.

**Leadership Training Represents a Unique Evaluation Challenge**

Leadership training in many ways is the face of the corporate training function for senior
executives. It represents a highly visible area of investment in training dollars, with Bersin
estimating that $14 billion US dollars are allocated toward these activities in the U.S. alone
(Loew and Leonard, 2012). In a joint report by McKinsey & Company and the Conference
Board (2012), it was noted that when over 500 executives were asked to rank their top three
human-capital priorities, leadership development was included as both a current and a future
priority, and nearly two-thirds of the respondents cited leadership development as their number-
one concern.

**Transfer of Training**

Not all barriers to effective evaluation have to do with the practitioners themselves, but
are closely related to the nature of the training being evaluated and the organizational context in
which it occurs. The “transfer of training” has been defined as “the effective and continuing
application by trainees to their jobs, of the knowledge and skills gained in training – both on and
off the job.” (Broad and Newstrom, 1992, p. 6). In other words, the learning that takes place in
the classroom needs to be transferred back into the work setting. In Baldwin and Ford’s (1988)
meta-analysis of transfer of training literature, they pessimistically concluded that not more than
“10 per cent of these expenditures actually result in transfer to the job” (Baldwin and Ford, 1988,
p.63). More recently and positively, the Learning and Development (2008) Roundtable Learning
Effectiveness Survey cited earlier estimated that the top 20% of programs they evaluated demonstrated learning improvement of 42%, compared with 13% for the bottom fifth, with the average learner improvement of 35%. The same study concluded that 90% of the variation in application is a function of motivation to apply. Those programs that achieved the highest application rates improved learner performance and business performance 50% and 28% more than the least effective programs, respectively.

Not all training is the same, and therefore the metrics, evaluation methods, and criteria must be tightly attuned to the different types of training. For example, sales training lends itself more easily to measurement, as the number or amount of sales can be captured as a dependent variable and contrasted either with pre-training levels or against a control group of individuals who did not attend the training.

Leadership training, by contrast, represents a more complex evaluation challenge. What constitutes effective leadership depends a great deal on context in the absence of a clear dependent variable. How might a successful outcome be defined? How will it be isolated? Who will report it - the individual, the individual’s manager, or peers? Is there a maturation factor required before results can be observed? Is it the same for everyone? Further, leadership training is rarely able to replicate the environment in which the learning is meant to be applied. It is more likely that in the classroom, the participant will learn a broad principle that the leader can potentially apply in multiple situations and in slightly different ways.

Cherniss and Goleman (1998) highlight a number of these challenges in their technical report “Bringing Emotional Intelligence to the Workplace.” Cherniss and Goleman draw the important distinction between cognitive and emotional learning, noting that the latter is more challenging as it frequently involves the need of the learner to come to terms with deeply
engrained habits that have been forged over a lifetime. As Goleman (1995, 1998) has established, leadership effectiveness has been more closely correlated with emotional intelligence than cognitive intelligence, particularly at more senior levels. It is axiomatic that if one lacks self-awareness, it will be difficult to be aware of others’ needs, and a leader who is unable to calibrate emotionally with others will have a difficult time in aspects of leadership that require empathy, such as coaching, delivering feedback, and motivating others. As a result, leadership training that does not account for learner motivation at a social and emotional level is not likely to be very effective in leading to changed behaviors and application on the job.

Similarly, effective evaluation must be able to account for changes that lead to more effective application of these so-called soft skills.

Yelon and Ford (1999) introduce a helpful distinction when thinking about training transfer in what they have referred to as closed and open skills transfer, respectively. Closed skills transfer refers to a narrowly applicable skill, e.g., the application of new knowledge of a phone system to the identical phone system when returning to the office, which leads to sales calls. Open skills transfer refers to learning related to generalizable concepts, rules and principles.

Furthermore, leadership development programs generally focus on skills that require time to develop in order to become more effective in practice. This means that the timing of the evaluation is also an important consideration to obtain an accurate measure of what has been applied and the impact that the actions have produced.
DISSERTATION: APPLIED META-EVALUATION

Brinkerhoff’s Success Case Method as a Promising Approach for Evaluating Leadership Development Programs

Despite the fact that the training evaluation field remains dominated by Kirkpatrick’s 50-year old framework, research specifically into the area of transfer of training (Blume, Ford, Baldwin and Huang, 2010; Wickens, Hutchins, Carolan and Cummings, 2011) and impact evaluation is creating a sea of change that represents an important step forward. Within this context, the Success Case Method (“SCM”) developed by Robert Brinkerhoff (Brinkerhoff, 2005) has shown promise as an approach to assess the impact of corporate leadership training. The SCM is essentially an approach to identify “success cases” (individuals who are effectively applying the training) and analyze the business impact, highlighting for stakeholders those factors that enhanced or limited this application.

Brinkerhoff himself has a rare career that has straddled public education as faculty (now emeritus) at Western Michigan University and as a consultant to corporate training. One of the distinctive elements of his approach is that it takes just as seriously those organizational factors that would either enhance or inhibit the transfer of training (e.g., the role of the participants’ manager) as well as a simple focus on the training. While Kirkpatrick’s levels tend to be almost universally utilized, the joint ASTD/i4cp industry evaluation study cited earlier notes that nearly half of study’s respondents reported using the “Brinkerhoff” or “Success Case Method” (SCM) to evaluate training (Bingham, 2010). This approach will be described in greater detail in Chapter II.

Meta-Evaluation as a Tool to Bridge the Gap Between Research and Practice

Meta-evaluation is a term that was first introduced in 1969 by Michael Scriven to describe “any evaluation of an evaluation, evaluation system or evaluation device” (Scriven,
Meta-evaluation can be thought of as a safeguard for consumers of evaluation studies, as it can help them to decide to the degree, which they can accept the conclusions reached in a given evaluation, whether it is of a product, program or service. In his article, “The Meta-Evaluation Imperative,” Stufflebeam (2001) refers to the process as a “professional obligation of evaluators” and provides a more operational definition for the process. Specifically, he defines it as:

The process of delineating, obtaining, and applying descriptive information and judgmental information—about the utility, feasibility, propriety, and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness, and social responsibility—to guide the evaluation and/or report its strengths and weaknesses (Stufflebeam, 2001, p. 185).

In considering some of the potential evaluation flaws that might be identified and remedied through rigorous and systematic meta-evaluation, Stufflebeam’s list (2000) below is instructive:

1. Inappropriate criteria;
2. Biased findings;
3. Technical errors;
4. Unjustified conclusions;
5. Ambiguous findings;
6. Unwarranted recommendations;
7. Excessive costs;
8. Inadequate interpretation to users; and,
9. Counter-productive interference in the programs being evaluated.
Given that poor evaluations can lead to misguided decisions with adverse consequences, training professionals bear an ethical responsibility to conduct or sponsor meta-evaluations to ensure that any decisions being reached are based on solid evaluation evidence. It is a question of professional accountability to “evaluation consumers” and other stakeholders potentially impacted by the program or program evaluation. Stufflebeam and Coryn (2014), argue that meta-evaluations “should be grounded in sound standards for evaluations, conducted formatively to guide and ensure the quality of evaluations, and conducted summatively to judge the evaluation at hand in terms of such factors as utility, feasibility, propriety, accuracy and accountability.” (p. xxxi). While meta-evaluation generally refers to evaluations of specific program evaluations, the same meta-evaluation principles can be applied to assess a given evaluation method or approach.

For corporate learning programs, there are several goals that a meta-evaluation would hope to achieve, including to: 1) investigate how the evaluation (or evaluation method) is implemented; 2) examine how the evaluation (or method) can be improved; 3) determine how worthwhile the evaluation (or evaluation method) is to their stakeholders; and 4) measure how the costs direct, indirect, and opportunity, compare to the derived benefits.

Which Standards Should Guide a Meta-Evaluation?

The Program Evaluation Standards (Joint Committee Program Evaluation Standards, 1981, 1994, and most recently in 2011) articulate 30 evaluation standards against which a program is evaluated. These 30 standards have been organized under four criteria: utility, feasibility, propriety, and accuracy. Utility refers to the usefulness or ability of the evaluation to serve the information needs of the intended users. Additionally, Scriven (2010) makes a helpful clarification that utility should not be confused with utilization. Utility is the property that maximizes utilization; utilization, being what the organization does with the evaluation,
ultimately outside the control of the evaluator. The concept of credibility is also important. Credibility forms a critical but not sufficient condition for utility. If an evaluation is not credible, it will not have much utility. On the other hand, an evaluation may be credible but still potentially missing the mark relative to the consumer’s expectations. If paraphrased into layman’s terms, the consumer might say “that’s interesting, and I trust that it is right, but it is not that useful to me.”

Feasibility refers to assurance that the evaluation is practical, viable and cost effective. The methodology used and processes implemented take into account the organizational context and must be carried out in ways that do not disrupt organizational routines or be viewed as overly intrusive in either the time commitments or access to information required to complete the study.

Propriety refers to the legality, proper ethics, and regard for the well-being of both the individuals who participate and the stakeholders who will be impacted by the results. The ethics, principles, and ideals go beyond the domain of evaluation and also reflect any other relevant professional, legal, moral or contractual agreements that could potentially be invoked by the study.

Accuracy relates to the standards that are meant to ensure that the evaluation will reveal and communicate defensible information, lead to justifiable conclusions, and deliver impartial reporting of findings. Accuracy also is concerned with the extent to which an evaluation is truthful or valid in the scope of what it communicates in consideration of the different components of the program being evaluated.

The standards and criteria have been endorsed by the American Evaluation Association (Sanders, 1994) and have also been delineated by Maher in “The Resource Guide for Planning
and Evaluating Human Service Programs” (2000). As testament to the growing emphasis in the evaluation field of the need to ensure that evaluations themselves are systematically reviewed and held accountable, the Third Edition of the Program Evaluation Standards published in 2011 (Yarbrough, et al., 2011), included an additional fifth criteria—accountability.

Operationally, accountability refers to the “responsible use of resources to produce value” (Yarbrough, et al., 2011, p. 226). The standards are explicit in their call for all evaluations to be systematically meta-evaluated. These Evaluation Accountability standards require:

1. Evaluation documentation. Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.

2. Internal meta-evaluation. Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.

3. External meta-evaluation. Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external meta-evaluations using these and other applicable standards.

The standards and criteria used to conduct a meta-evaluation should be based on the four main evaluation criteria (utility, feasibility, propriety, accuracy) and the applicable standards that support them. That said, meta-evaluations vary greatly in terms of scope, complexity, duration, available resources, environment and intended use. Depending on the purpose and rationale of a given meta-evaluation and the particular needs of its sponsors, not all of the standards need to be applied to constitute an effective meta-evaluation. Rather, the meta-evaluation should consist of those standards and criteria that are deemed as best able to cast light on the key features of evaluation quality that are to be assessed. What is most important is that “a meta-evaluative
In summary, while corporate training has received more focus and investment in recent years, its ability to be an integrated strategic partner to the business has been compromised by an inability to both deliver and credibly demonstrate its value. In order to remedy this precarious position, corporate training must first address its capabilities and practices in metrics and evaluation.

This integration failure represents a gap between research and practice, as both the quantity of research and number of practitioners have grown exponentially and outpaced what has been put into practice over the past 50 years. This gap is due to multiple factors. First, corporate training evaluation practices have evolved in parallel, but largely independently of, the broader evaluation field. Second, as noted earlier (see Kraiger, McLinden and Casper, 2004; Moller and Mallin, 1996; Learning and Development Roundtable, 2009; Swanson, 2005;), there is good evidence to suggest that corporate training practitioners frequently do not have an orientation toward evaluation and metrics, the ability to identify different evaluation methods or knowledge of obtained from relevant academic research. Third, evaluation has been viewed primarily as an after-thought with few resources being allocated toward evaluation. Fourth, the nature of leadership development programs present unique challenges for defensible and credible metrics. Lastly, few corporate training functions have a robust metrics and evaluation strategy to ensure evaluation approaches are being deployed appropriately, with key metrics being tracked in a consistent and sustainable way.

Meta-evaluation, or evaluation of evaluations, is emerging now as a core component of responsible evaluation, yet it does not appear in any of the literature to be a component of any
DISSEMINATION: APPLIED META-EVALUATION

corporate training evaluation approaches. Meta-evaluation, therefore, represents an opportunity
for corporate training functions looking to increase their evaluation capability and effectiveness
and bridge the gap between evaluation research and practice. If deployed judiciously,
systematically, and consistently, meta-evaluation holds promise as a tool for those organizations
looking to incorporate multiple evaluation approaches and metrics in a more strategic and
holistic way.

The Study

This study applies meta-evaluation standards to conduct an internal meta-evaluation of an
evaluation of a global leadership development program that was launched in a global bank. The
_evaluand_ (i.e., the subject of the evaluation) is an evaluation that had been produced using
Brinkerhoff’s Success Case Method (SCM). This approach was selected by the host
organization because the program management team viewed the SCM as a promising method
that might improve the organization’s ability to assess the impact of the leadership development
program.

As the head of the department that designed and deployed the leadership program as well
as sponsored the SCM evaluation, the researcher is interested in understanding the extent to
which the SCM approach met its intended purposes. Given the researcher’s involvement with the
program and SCM, he played a dual role as participant-observer in this investigation. The
strengths and limitations of this approach will be elaborated on further in the Discussion chapter.

While some treatment will be given to the results in relation to each of the evaluation
standards and related research questions, the discussion will focus primarily on the extent to
which the meta-evaluation was able to determine the utility and feasibility of the Success Case
Evaluation. This is not to say that accuracy and propriety are less important to a valuable
evaluation; on the contrary, they are foundational. If one cannot trust the accuracy of the findings or the method that produced them, the evaluation has little value. Similarly, if the method or study itself introduces any biases which call into question the representative nature or propriety of the study, these also threaten and undermine the value of the study. That said, these two standards can be evaluated independently and are less dependent upon the collective perception of key stakeholders. On the other hand, feasibility and utility represent two of the primary hurdles an evaluation must clear if an evaluation is to be as having been viewed as worthwhile. The extent to which a given evaluation is able to ‘clear these hurdles’ is highly dependent on the largely subjective perception of stakeholders relative to the cost/benefit of the evaluation, whether that is in the effort required to conduct the study “feasibility” or the “utility” of the evaluation to the organization.

**Research Questions**

The meta-evaluation (the “study”), in the first instance, attempts to answer the following four questions corresponding to each of the program evaluation standards.

1. To what extent did the SCM meet the evaluation standards for utility in the context of the organization in which it was conducted?
2. To what extent did the SCM provide a feasible approach for the organization in which it was conducted?
3. To what extent did the SCM meet the evaluation standards related to propriety?
4. To what extent did the SCM meet the evaluation standards for accuracy?

The final two research questions are aimed at gaining an overall understanding of the extent to which the SCM was successful and the meta-evaluation itself was useful to the organization.
5. To what extent did the Success Case evaluation succeed in determining the impact of the leadership development program?

6. To what extent was the meta-evaluation useful to the organization as a means to determine the efficacy of the Success Case Method for the leadership development program?
Chapter II: Background of the Organization, Leadership Program and Success Case

Method

The Bank

The host institution for the study is a global financial institution that conducts business in over 100 countries. The organization was hit hard by the global financial crisis and over a period of three years was forced to reduce its headcount by a third. Prior to the financial crisis, the bank had operated in a highly decentralized and siloed manner, which led to duplicative efforts and inefficiencies. During the financial crisis, the company began to accelerate efforts to achieve economies of scale through centralization of its global functions (e.g., legal, human resources, finance) and global product lines.

In many ways, the Learning and Development function mirrored the rest of the organization. For example, in 2005 when the researcher joined the company, there were approximately 30 separate departments designing and developing leadership content along business and geographic lines. At that time, Executive Development was the only part of the L&D function that offered training across all business, function and geographic lines, focusing on senior leadership population (e.g., Managing Directors). From 2006 to 2008, the Leadership and Executive Development departments were combined to form a Center of Excellence (“COE”); this initiative centralized all design and delivery of leadership development offerings into a single work group. The COE became responsible for the design of the leadership development core curriculum and the content would be delivered in each of the company’s four geographic regions: North America (“NAM”), Latin America (“LatAm”), Europe, Middle East, and Africa (“EMEA”), and Asia Pacific (“APAC”).
In order to ensure the relevancy of content across a diverse population of businesses, geographies and functions, the COE adopted a leadership framework it hoped would transcend the unique aspects of any particular business to hone in on the unique requirements of key roles at different levels of the managerial hierarchy. The Leadership Pipeline model, as described by Ram Charam, Stephen Drotter, and Jim Noel, provided the basic foundation for this leadership framework (Charam, Drotter, and Noel, 2001). The leadership pipeline concept proved to be valuable at highlighting the importance of transitions from one type of role to another (i.e. from individual contributor to manager of others) and also outlined the shifts in skills, priorities, and time management that would be required for success at each level. The model also highlighted the paradox that the very skills that had made an individual successful in one role could become obstacles to success in the next.

The bank simplified the Charam, Drotter and Noel (2001) Leadership Pipeline model from six levels to four levels. These levels are: (a) manager of self (individual contributor); (b) manager of others; (c) manager of managers; and (d) executive.

**Genesis of the Leadership Program**

As part of the overall implementation of the Leadership Pipeline model, the COE took an inventory of the existing programs that were targeting the group of senior managers a level below the executive population- the Senior Vice President and Director levels. Upon review, it was clear that there was no existing training program that adequately served this population nor addressed the unique challenges facing these managers of managers.

The existing information relative to the strengths and areas of opportunity for this level of the population was consistent with the aggregate data collected at the executive level, which included themes from 360-degree feedback, development plans, and company culture surveys. The key gaps identified coalesced around people management skills, including delivering
regular, constructive feedback, developing successors for key positions, coaching, and managing performance. These gaps had previously been identified in the Managing Director population, which in 2006 led to the creation of a three-day residential training program for individuals identified as high potentials.

With the core elements for a strong program already present at the Managing Director level, work was undertaken to ensure that the content was appropriate for this lower management level. The Leadership Pipeline model was examined to determine which skills would be most appropriate to be addressed within the context of the new program. Additionally, a conviction had emerged within the new COE that leadership development programs play an important role in communicating company expectations and in the development of a common culture. As a result, it was decided that company employees serve as the primary faculty of the program versus external facilitators. With this in mind, the original external vendor-partners were re-engaged to assist with the re-design of the program and subsequent faculty development.

The program was piloted in North America in 2008 and then delivered in each region throughout 2009 with a “train-the-trainer” element designed to develop future facilitators and coaches for the program. The Director of Executive Development (the researcher) and a facilitator from the vendor partner co-delivered these initial programs and “certified” new faculty in each of the company’s regions. The program became the first of four programs of the core curriculum that would ultimately be launched.

**Description of the Training Program**

**Target population.** The program was a three-day residential program that targeted senior managers of managers. *Senior* has been operationally defined as those managers who have been in a “manager of managers” role for at least two years. Participants were nominated to attend the program by their businesses and had at least five direct reports, been in role for at least two
years, and considered as having the potential to take on larger leadership roles in the future.

Seats were allocated to the businesses and functions based on their relative size within each respective region, with a view that participants had the opportunity to interact with a representative mix of colleagues from different areas while at the program. Cohorts were generally limited to 35 and participants were assigned to tables with 6-7 colleagues with whom they primarily worked throughout the duration of the program. There was also an internal coach that was assigned to each table.

Coaches. The coaches in the program were primarily seasoned human resource professionals who have been “certified” in the bank’s coaching framework. The coaches fulfilled the role in addition to their primary responsibilities. To be considered for the role, potential coaches were first nominated by their respective Senior Human Resources Officer (SHRO), and then underwent an interview process with a committee of senior coaches. The coach’s role was not only to help ensure that table exercises ran smoothly, but also to conduct two short (30-minute) coaching sessions with each individual throughout the three day experiences. As part of these individual sessions, coaches focused on helping participants to understand their individual assessment data and formulate an “action plan” to address the leadership challenge each participant brought to the program as part of the pre-work. Coaching continued after the program with the intact table group for six months with group calls scheduled at one, three and six months. Program participants also had the opportunity to participate in ad hoc individual sessions with the coach depending on the coach’s availability.

Faculty. There were generally two faculty assigned to deliver each program. The selection process for faculty was similar to the process for coach selection. Faculty was also from Human Resources, with most having been members of the senior L&D and Talent functions. These faculty were handpicked by the program management and validated with the senior human
resource officers (SHROs) of their respective businesses or functions prior to participating in the train-the-trainer sessions.

**Program objectives.** The program objectives were formulated based on three key inputs: (a) the requirements of the manager of managers’ role as described in the company’s Leadership Pipeline model; (b) skill gaps identified in a review of aggregate 360 data from the Executive Development high potential program; and (c) consultation with senior business leaders.

The stated program objectives were as follows:

1) Move from a day-to-day focus (managing) to a future focus (leading) by providing participants with tools to more effectively:
   a. Develop self as a leader. Assess personal leadership capabilities; manage risk factors and potential derailers. Build and act upon personal development plan.
   b. Develop a coaching strategy. Actively develop and hold direct reports accountable for coaching, motivating, and managing their team members.
   c. Lead a high-performance team. Assess and manage (your) teams’ capabilities, business performance, and productivity.
   d. Foster and leverage relationship. Demonstrate the organizational savvy to navigate in the company’s highly complex, matrix structure. Utilize informal networks to work effectively across boundaries.
   e. Inspire others to action. Articulate the vision and direction for (your) team to align with business strategy. Build and refine your leadership style and presence to influence and motivate others.

2) Create an opportunity to assess (your) current leadership style and plan one or two shifts that will increase (your) effectiveness.

3) Prepare (you) to transfer key selected insights to your direct reports after the session.
Agenda overview. The program was designed to be a three-day residential experience, with the expectation being that participants would stay at the hotel or conference center in the evenings in order to build relationships with colleagues. The overarching progression of the program moves from focus on oneself, to one’s team, to the broader organizational network. Day 1 focused on one’s self as a leader. Day 2 focused on the team or teams under the leader’s purview. Day 3 focused on expanding one’s influence to lead across boundaries. A brief description of the activities, exercises and tools associated with each of the program components follows.

Pre-Work

In advance of the program, there are three requirements that the participants must complete.

The bank’s proprietary Manager of Managers 360-degree assessment. Managers were first administered a 360-degree survey by email. This survey is a proprietary multi-rater tool developed to highlight the company’s Leadership Standards as they relate to the expectations and requirements of Manager of Managers. The tool consists of 52 items and takes approximately 30 minutes to complete. In addition to answering each of the 52 questions for themselves, participants receive input from their manager(s), peers, direct reports, and a category known as “others.” The overall purpose of the tool is to give participants multiple perspectives on how their leadership is perceived and help them gain greater self-awareness around any gaps in their own perceptions versus those of the stakeholders completing the assessment.

Hogan Development Survey. In addition to the 360-degree feedback survey, participants completed a personality assessment known as the Hogan Development Survey. This assessment was designed by Robert and Joyce Hogan as part of a suite of three assessments aimed at helping individuals understand how their personalities influence performance and satisfaction in the
workplace (Hogan and Hogan, 1997). The Hogan Development Survey specifies 11 dimensions of personality that are believed to be predictors of potential career ‘derailment.’ The fundamental premise of this tool is that each individual possesses relatively static predispositions inherited at birth and solidified during early childhood. These predispositions have the potential to be activated and thus manifest themselves negatively when an individual is under stress. When activated, a particular behavioral tendency, which may be functional or even desirable at a moderate level, has the potential to be amplified into behaviors that could negatively impact the individual’s performance. Self-confidence, for example, is generally a desirable characteristic; however, when self-confidence is amplified under stress, it can cross over into behaviors that could be perceived as arrogance.

The Hogan Development Survey aims to create awareness in individuals of those predispositions which might place their effectiveness or careers at risk. While these predispositions are viewed as a fixed part of personality that cannot be altered, the hope is that through greater awareness, they can be managed. Participants are encouraged to cultivate an awareness of those internal and environmental triggers that induce stress in order to develop strategies to avoid de-railing-type behaviors.

In the Leadership Program, participants were encouraged to reflect on their leadership effectiveness by reviewing the 360-degree feedback report side-by-side with the Hogan Development Survey. Specifically, participants were advised to look for evidence of their derailers in their 360-degree feedback. If any potential derailers appeared to be manifest in the feedback, the participant was encouraged to understand the cues that activate the behavior and develop strategies to keep this behavior in check. If derailers do not appear, it does not necessarily mean that they are not present. It may mean that the individual has not had exposure to the triggers that might activate a given derailer, or that the individual has already learned
successful self-management strategies. It is not uncommon with mid-career executives to observe that the highest propensity derailers are those that are best managed (least manifest) because the individual has likely learned from experience. Instead, the discovery of secondary derailer tendencies, which are less obvious, may provide some of the more illuminating insights.

**Personal leadership challenge.** For the third component of the pre-work, participants received instructions to prepare a written “personal leadership challenge.” This “personal leadership challenge” is meant to reflect a current high priority work situation or goal that requires more effective leadership from the individual to successfully meet the challenge.

This pre-work component was designed to accomplish a number of goals. First, it positions the upcoming course as practical and relevant for the day-to-day work versus a theoretical exercise. Second, preparing the challenge helped participants to mentally prepare and anticipate areas of their own leadership where they would like to focus. Lastly, by putting the challenge into writing, participants were better prepared to share the challenge in a clear and succinct way with colleagues while at the program. The personal leadership challenge forms the “raw material” which participants take to the program and use as a foundation for building their action plan by the end of the experience.

**The Program**

Participants were seated at tables of approximately six to eight colleagues with a table coach. Seating was designed to be “max-mix,” with a maximum diversity of participants representing businesses, functions and gender. Participants were given a workbook and a worksheet known as the Personal Planning Template. The overall experience is highly interactive and reflects a mix of classroom lecture, open discussion, table exercises, role-playing, peer coaching, and individualized coaching. As participants progressed through the different modules, they were instructed to add insights and data into their personal planning template,
which would then culminate in two to three actions that they are then committed to taking at the program’s end (i.e., their personal action plan). Participants remained at the same table throughout the experience and were invited to participate in group coaching sessions with their table colleagues at one month, three months and six months after the training experience ended.

The Program’s Deployment and Evaluation

After piloting the program in 2008, the program was formally launched as a global offering in 2009. For the inaugural session in each region, the researcher and one external faculty delivered the program with potential future internal faculty in attendance. In the two days following the program, the new faculty attended a ‘train-the-trainer’ session to become “certified” to deliver programs in their respective regions. In total, the program was delivered 10 times to a total of 296 participants globally in the first year. At the conclusion of the 2009 delivery cycle, Executive Development (headed up by the researcher) contracted two external evaluators to conduct a Success Case evaluation of the program to better understand its impact within the organization.

The SCM was selected because it was viewed as a promising approach that could help the organization to better quantify the impact of its programmatic investments in the leadership training in the future. The external evaluators for the Success Case approach were selected by virtue of their close affiliation with Dr. Rob Brinkerhoff (developer of the SCM) and for their broad experience in applying the SCM within corporate settings. At the time of the study, the evaluators had Dr. Brinkerhoff’s authorization to utilize, market, and certify others in the Success Case methodology. The evaluation took place between February 2010 and May 2010, with 200 of the 296 program alumni participating. At the end of the evaluation, the evaluators produced two reports. One report contained highlights of the study that could be shared with a broad set of stakeholders; the second report contained more detailed information on every aspect.
of the evaluation and data that were collected and were intended primarily for the internal team that had sponsored the evaluation.

The Success Case Method

**Background and process.** The SCM is an approach developed by Dr. Robert Brinkerhoff where the evaluator deliberately seeks out success cases which highlight instances where the learning from the program is being applied by participants with impact. The evaluator then compares and contrasts these success cases with non-success cases in order to improve the program and “tell training’s story” (Brinkerhoff, 2005). According to Stufflebeam and Corwyn (2014), one of the benefits of the SCM is that it is a “relatively quick yet defensible means of gathering critically important information for use in program improvement. The approach may be employed in conclusion-oriented summative evaluations, but mainly it is intended for use in formative evaluations aimed at program improvement” (p. 137)

The SCM is the byproduct of Brinkerhoff’s unique experience as both academic researcher and faculty at the University of Western Michigan and as an evaluator of training programs in the private sector. For Brinkerhoff, the key question in evaluation is not related to how the measurement occurs, but why it should be done in the first place. He argues that if organizations and practitioners can reach some clarity around the *why* of evaluation, the *how* begins to take care of itself. For example, if training and development is viewed as a staff benefit, the *why* question worth answering is, “Why do the participants appreciate and value the experience?” If, on the other hand, training is viewed as a business driver, the key questions no longer focus on level of participation or even how much learning has taken place, but on how

---

5 This question roughly corresponds to Kirkpatrick’s Level One (Reaction)
and to what extent learning is being applied back on the job and what results are being achieved that can be attributed to the program.

As the evaluation moves to application and results, it also needs to shift its focus from the program’s quality, design, and facilitation to the learner and organizations’ performance system. For Brinkerhoff, the traditional training-focused evaluation strategy (vs. performance-focused) poses several significant risks. First, in addition to ignoring the performance system factors that influence training impact, it unintentionally undermines the performance partnership with line management by obscuring the overall ownership for performance, which ought to lie primarily with management. Responsibility for application, if it is indeed important, should lie with the learners, their managers, and senior leadership (those who establish policies, incentive structures, and sponsor the programs, etc.), and not the training function. Similarly, when things go well, training functions often receive the credit, which underestimates the key contributions that have been made throughout the system to make it conducive for achieving the results.

A second risk stemming from a training-focused evaluation model is that it deprives managers of essential feedback that would enable them to play a more productive role in setting the stage for and reinforcing the application of the learning by their direct reports (i.e., the participants). Brinkerhoff does not deny the need for training organizations and practitioners to be held accountable for conducting a thorough needs assessment, ensuring high quality delivery of the program, and for reporting basic aspects of the training. Those are simply table stakes. More importantly however, Brinkerhoff argues that training practitioners must keep the broader context in focus and utilize evaluation to elucidate insights from the surrounding performance system, particularly those that could potentially increase the likelihood of successful application of the learning in the workplace.
Brinkerhoff’s (2006) inspiration for the development of the SCM emanated from two insights, which have become axiomatic for the SCM. The first insight is that training programs generally produce predictable results. The second insight is that all training takes place within in a specific context and complex system.

**Predictable results from training.** According to Brinkerhoff, all training programs produce three basic categories of results. The first category represents a small minority of participants who actively use the learning and achieve concrete and valuable results. The second category consists of another small minority of the participants who fail to use the learning at all once they return to their day-to-day work. The third and largest category represents the many that utilize some of the learning but apply it inconsistently and do not achieve any noticeable results that could be directly attributed to having participated in the training.

This reality, while making intuitive sense, is often ignored when traditional statistical methods are applied to evaluation, producing what Brinkerhoff refers to as a “tyranny of the mean,” (Brinkerhoff, 2005) in which a focus on the mean obscures the actual impact of the training and reduces the opportunity for learning that can be gleaned from studying those groups at the extreme ends of the distribution curve. For example, if only the average level of successful learning transfer following attendance at a program is addressed, one might conclude that the training was mediocre. A more likely reality, however, is that there were some individuals who achieved outsized results attributable to the program and there were others who simply achieved nothing. Only focusing on the mean would fail to recognize either of these groups, and the mean drives away focus from these groups that reveal the most about the training. The SCM attempts to address this oversight by inquiring more deeply into the highest and lowest impacts to be able to better understand the mechanisms that could be most instructive
DISSERTATION: APPLIED META-EVALUATION

for increasing future impact. Not only does this focus highlight more accurately where success is happening, it subtly begins to shift the onus of the evaluation responsibility from the program and the training organization over to the learner and the performance environment in which the learner operates. This of course presumes that the training design is already closely aligned with the requirements of the business.

**Training takes place within an organizational context.** The second reality that Brinkerhoff addresses with the SCM is that training never operates in isolation. Rather, training is embedded into a specific context and complex system. Brinkerhoff estimates, as a general rule, that approximately 80% of impact is attributable to the training context, whereas only 20% can be attributed to the training itself. This learning context consists of those factors outside of the classroom that impact the likelihood that transfer of training will occur (e.g., resources, individual capability, motivators, managerial support, opportunity, accountability and incentives for follow through). Paradoxically, Mooney and Brinkerhoff (2008) estimate that the before and after elements of the learning process receive only 10% and 5% respectively of the overall time invested by those managing the training, with close to 85% of effort focused on the learning event itself. Therefore, if one of the key goals of the evaluation process is to maximize training transfer or impact, the focus needs to be more holistic and not limited solely to the delivered program. Brinkerhoff believes that by expanding the area of focus, a clearer understanding of the mechanisms and components that either enable or hinder training transfer will emerge and provide guidance for how to maximize impact.

**How SCM addresses and leverages these insights.** The SCM consists of two primary phases, broken down into a total of five steps, which are designed to deal with the two realities described above.
In the first phase, a survey is administered in order to provide a quantitative backbone to the evaluation. The survey determines the general distribution of which program participants are applying their learning and receiving worthwhile results, and those who are not. The first phase serves two purposes. The first purpose is to identify the most successful and least successful users of the training. The second purpose is to gauge the breadth of application and positive results (impact) among the participant population.

The second phase involves a series of in-depth interviews with a subset of the participants at the two extremes of the distribution curve. In these interviews, evaluators ask questions aimed at identifying and understanding key aspects of the training and the “performance system” that differentiated the successes (success cases) from the non-successes. In practice, the responses generated are frequently opposite sides of the same coin. For example, those who experience a high degree of success in post-classroom impact often report that their direct manager inquired about the learning and regularly followed up on progress. On the other hand, many who state they have not applied the learning back on the job indicate that their managers had little or no interaction with them relative to the program or applying the learning. The interviews serve as a helpful methodological check to eliminate any false positives identified through the survey, where those reporting successful application cannot truly attribute these actions directly to having attended the program. The interviews also provide rich information relative to the ways that the learning is being applied, the barriers that are being encountered, and help surface specific strategies the learners employ.

When taken together, the two phases enable evaluators to develop a clear and more nuanced picture of the overall impact of the training while extrapolating potential value that could be realized if more were to effectively apply the training. By studying both the context and
those factors that either enable or hinder successful application/transfer, the evaluator is able to make recommendations to future trainees, key stakeholders, and those who manage the program about how to increase impact.

After the analysis, the Success Case Evaluation attempts to answer eight questions. The conclusions that are drawn as a result of answering these questions address fundamental issues that key stakeholders might have in relation to the program in focus:

1. What, if any, impact was achieved? This question is usually foundational for the evaluation and the report. In addition to highlighting the overall impact, it also provides examples of best-case outcomes as an illustration of a successfully applied training.

2. How widespread is success? In this case, the survey that was administered enables the report to communicate the overall percentages of participants who were not successful, moderately successful, or highly successful in applying the training.

3. Did the training work better in some parts of the organization or with some kinds of participants compared to other parts or with other people? To what extent are the participants’ backgrounds or roles factors that could explain success of the training in specific parts of the organization?

4. Were some parts of the training more successfully applied than others? Not all training within a program will be equally applicable, given the varying situations and backgrounds of the different learners. It is however valuable to understand which tools, techniques, models or applications tend to be used more or less than the others.

5. What systemic factors were associated with success and a lack of success? By looking at the overall performance system, this question seeks to explain why the training was more
successful in one area versus another and identifies factors that could be more closely linked to success or lack thereof.

6. What is the monetary value of the outcomes produced? Where quantifiable, the SCM finds it desirable (but not necessary) if actual amounts of money can be tied to the application of the learning (e.g., where a program led to actions that either increased revenue or reduced expenditures).

7. What is the unrealized impact potential of the training? Given that no training is one hundred percent successful, this question seeks to draw conclusions about the opportunity cost of not applying the training as well as extrapolate what value could have been generated had more people applied their training successfully. Discovering unrealized impact potential is one of the strengths of SCM, and will be explored more fully in a future section, as the ability to “tell training’s story,” including to highlight the “value left on the table” helps stakeholders understand how different actions might lead to more positive outcomes.

8. How do the benefits of the training compare with costs? Three related calculations help answer this question: the benefits-to-costs ratio (“BCR”), return on investment (“ROI”) and payback period (“PP”). Depending on the company that sponsors the training and evaluation, any one of these measures may be preferable. In practice, the challenge of applying these calculations has to do with both isolating the training benefits and converting measures into monetary value. To address the challenge of isolating and

---

6 The BCR is generally reported in statements such as, “For every $1 spent on training, $1.30 is returned.” ROI is a close relative to BCR, but is calculated by dividing the net benefits of the training by costs and then multiplying by 100 to convert it into a percentage. Using the same example above, the ROI would be 30%: ((1.30/1.00)*100) = 30%. The Payback Period is calculated by dividing costs by benefits and then reported in months or years. In the example above, this would be calculated by taking 1.00 (the cost) divided by 1.30 (the benefits) to obtain 0.77, which represents time in years. When multiplied by 12, (0.77*12) = 9.23, we obtain the equivalent in months, or roughly 9 months and 1 week to payback.
quantifying the training benefit, Phillips and Phillips (2010) have proposed four steps: (a) know the financial benefits of the program (e.g. increased revenue, cost savings, reduced turnover, etc.); (b) normalize the benefits so comparisons can be made; (c) assign quantitative meaning to a benefit; and then (d) make the ROI calculation. The key challenge in most cases is to first isolate the benefits that can truly be attributable to the program. Once isolated, these must in some way be quantified, which can be challenging, when some of the benefits are intangible (e.g., increased morale and engagement, time saved by employees, etc.). Despite these challenges, Phillips and Burkett (2008) have estimated that 80% of the measures that are important to an organization have already been converted to monetary values in most companies.

It should be noted that not all Success Case evaluation reports seek to address all eight questions, but rather are dictated based on the needs of the sponsoring organization. That said, most reports usually attempt to provide a holistic picture of what the impact of the training has been, those success cases where the learning has been most meaningfully applied, and a set of recommendations to increase the overall impact and the percentage of participants who apply the learning.

**Summary of the Success Case Evaluation Study of the Leadership Program**

In the case of the Success Case Evaluation of the leadership program (the evaluand), the evaluation attempted to provide answers to the following three questions:

1. What level of business impact has the training provided the company? (Is the company getting a positive ROI?)

2. Has the training enhanced manager of managers’ ability to move from a day-to-day focus (managing) to a future focus (leading)?
3. What improvements can be made to the program to achieve even greater levels of impact for future participants and for the company?

In the first phase of the study, an electronic survey was sent to the 296 participants who had attended the leadership program between March and December of 2009. The survey had a response rate of 65%, with a total of 194 of the 296 participants completing the survey. The second phase of the study consisted of participant interviews of 15 of the 22 randomly chosen impact survey respondents based on geography. It should be noted that this was a departure from the traditional SCM where, based on the survey results success cases and non-success cases are identified for interviews. Seven of the 22 were not able to schedule interviews for various reasons. These interviews were to validate and identify specific ways in which the participants had been applying the program and to assess the extent to which there was measurable business impact. Of these, four were identified as true “Success Cases,” while there were none where the participants stated they had not made any changes as a result of the program (“non-Success Cases”), although application varied by degrees. The third phase consisted of analysis of all the data and production of the evaluation report.

Success Case Evaluation Conclusions and Recommendations

What follows are a summary of the conclusions of the Success Case evaluation:

1) The leadership program has had an impact, although it seems to be more personal development than tangible business impact.

- While several participants (29%) reported using the training to produce solid business impact in their surveys, only 14% of the interviewees were able to provide tangible/measurable evidence of business impact.
Interviews did reveal that participants felt better about themselves as leaders and colleagues as a result of the program, even though they could not provide quantifiable business results.

Working cross-business provided great insight and personal benefit, but little quantifiable impact.

The most consistently reported use of the program was communicating better with their teams as well as using 360 feedback findings.

Even with strong influence of these two areas, few respondents could report business impact.

The face-to-face networking and contact with others was seen as the most valuable part of the program.

2) The organization is leaving considerable impact on the table from the program.

Significant results are being missed because many participants are only casually using the skills.

3) Managerial support of trainees is weak and not focused on business outcomes.

Few managers were actively setting expectations for participants before the program.

The quality and quantity of coaching from managers post-program ranged from strong to non-existent.

The reference to “these two areas” is somewhat unclear. It was the investigator’s assumption that the two areas refers to the previous bullet which highlights that participants consistently reported that they were communicating better with their teams and making use of what they had learned from their 360 reports.
Although there is more reported follow up from table facilitators than from managers, even the Table Facilitator follow-up process was inconsistent and varying in impact.

The following were the Success Case Evaluation recommendations:

1) Raise the expectations and accountability for leveraging the learning experience with senior stakeholders, managers, and participants.
   - The organization should expect greater business impact.
   - Participants should be expected to apply their learning to significant business opportunities/issues in their work units.
   - Embed “how I will use this training to drive business goals” in every segment of the program.

2) Increase manager engagement throughout the process.
   - Managers should be held accountable for supporting/coaching their participants
   - Executive Development should set standards and educate managers on their role

3) Consider suggestions from participants to refine content.

After reviewing the Success Case Evaluation Report with the evaluators, the internal team (Executive Development), worked along with the regional program manager and select program faculty and coaches to review participant feedback for program improvement, incorporate a summary of the Success Case findings into the program, and establish calls with participants’ managers in advance of the program. The final report was also shared with all faculty and coaches, the global Head of HR and the Chief Learning Officer.
CHAPTER III: Method

The Original Study (Success Case Evaluation)

The organization launched and delivered a leadership development program to 299 senior managers of managers globally in 2009. At the conclusion of the 2009 delivery cycle, the organization contracted two external evaluators to conduct a Success Case evaluation of the program to better understand its impact within the organization. The evaluation took place between March 2010 and July 2010, with 200 of the 296 program alumni still employed by the company participating.\(^8\)

The Present Study (Meta-Evaluation)

This meta-evaluation study evaluates the Success Case evaluation of the organization. This investigation may be classified as an exploratory case study for two reasons: (a) the small number of potential participants for this study who would have had the requisite exposure to both the program and its evaluation; and (b) this represents the organization’s first experience with meta-evaluation. Given these two reasons, a mixed-method approach was adopted to provide greater flexibility for the data collection. Data were collected via online questionnaire, semi-structured interviews, and through an archival review of existing documents. Process and goals for each aspect of data collection will be outlined in greater detail.

Meta-Evaluation Participants

The investigator personally invited 22 individuals to participate in the study. These individuals were regarded as important stakeholders to the leadership development program by virtue of the roles each had played at some level of the program’s design, delivery, or

---

\(^8\) Three of the 299 employees who had participated in the program were no longer employed by the bank at the time of the Success Case Evaluation
management. All participants were middle to senior-level managers in Human Resources and most were members of either the Learning or Talent organizations, including both the Chief Learning Officer and Head of Talent. Twenty-one of the 22 individuals recruited by the investigator participated in the study. Of the 21 subjects, 10 were male (including the investigator) and 11 were female.

The Executive Development area within the Learning function managed the leadership development program at the time of the Success Case evaluation. Consequently, the participants in the study who had the greatest exposure to the leadership development program were a female Vice President and the male investigator (author of this study), who was the Director of the Executive Development department. The Vice President was a direct report of the investigator and, in her role as Program Manager, had primary internal responsibility for working with the two external evaluators to conduct the original evaluation.

The original leadership program had been delivered unevenly across each of the company’s geographic regions. In order to retain consistency, efforts were made to mirror the proportion of participants in this study to be reflective of the program’s global deployment. Three and four participants were recruited from the Asia Pacific and Europe Middle East and Africa regions, respectively. One individual was selected to represent the Latin America/Mexico region. The remaining 13 participants in the study were based in the United States. The resulting mix of participants was proportionally reflective of the program’s regional diversity.

Informed Consent and Confidentiality

The investigator solicited participation in the study through direct contact and participation was finalized in writing via email. Prior to participation, all participants were required to sign and return an informed consent form, which was attached to the email invitation
DISSEMINATION: APPLIED META-EVALUATION
(see Appendix A). In addition, participants received a copy of the approval letter signed by the University IRB from the University for research involving human subjects. The informed consent form outlined the key elements of participation in the study, including: (a) a description of the Success Case evaluation report they would be expected to review; (b) the online survey; (c) the nature of the semi-structured interview (if applicable); and (d) the estimated time commitment required for each component.

Participants were informed of how their involvement in the study would improve the organization’s ability to measure the impact and ultimately increase the effectiveness of its training programs. Participants were advised that the intent of the research was to evaluate an evaluation of a training program, clearly stating that their individual opinions would not be the subject of any evaluation. Participants were promised a summary of the study’s findings upon completion. Lastly, participants were reminded that, while the study had been authorized by the company, their participation in the study was voluntary and not related to any company requirement, and that they could withdraw from participation in the study at any time, for any reason.

The informed consent letter also provided information regarding how the information collected would be utilized, who would have access, and the measures that would be taken to safeguard their personally identifiable information. Upon receipt of each participant’s signed informed consent forms, the investigator returned via email a counter-signed copy of the agreement.
DISSERTATION: APPLIED META-EVALUATION

Procedure

Data was collected via online questionnaire, semi-structured interviews, and a review of archival data (extant documentation relative to the program).

**Online questionnaire.** The investigator created a 16-item survey for online administration (see Appendix B) in order to obtain quantitative data regarding participant views relative to the utility, feasibility, accuracy and propriety of the Success Case evaluation. The investigator consulted the 27 standards from the 2011 Program Evaluation Standards (Yarborough, et. al, 2011) and extant meta-evaluation evaluation checklists (Stufflebeam, 1999; Stufflebeam, Goodyear, Marquart, and Johnson 2006) to guide the item construction. The resulting 16 items reflect those elements which, based on the investigator’s knowledge of the organization, were most relevant.

The 16 survey items were mapped to the four evaluation standards, associating three items with accuracy, four with feasibility, four with propriety, and five with utility. An independent third party reviewed, edited, and confirmed the survey and mapping. Each item of the survey began with the phrase “to what extent,” and responses were measured on a 5-point Likert scale, ranging from “not at all” to “to a very great extent.” Given the amount time that had elapsed since the original evaluation of the leadership program, and the fact that some participants would have had more involvement and line-of-sight into the original evaluation than others, an additional response option of “do not know/not applicable” was included for each survey question. Scoring of the items was as follows: 0 = do not know/not applicable; 1 = not at all; 2 = to a little extent; 3= to some extent; 4 = to a great extent; and 5 = to a very great extent. An open text box labeled “Comments” followed each item, enabling participants to elaborate on a particular response. Participants were provided a copy of the original Success Case evaluation
report for review prior to completing the survey in order to mitigate potential memory decay over the three-year gap between evaluation and meta-evaluation.

All 21 subjects completed the online survey. For the 16 participants employed by the bank at the time of this investigation, the organization administered the questionnaire using an internal company survey application. For the five participants who were no longer employees of the organization, independent third-party vendor SurveyMonkey® administered a password-protected online survey. After closing the surveys, the responses to the employee and non-employee surveys were combined into a single Excel spreadsheet.

**Semi-structured interview.** The collection and review of the survey data provided the substance for the semi-structured individual interviews. Eleven of the subjects were interviewed to elicit further qualitative information and to explore additional unanticipated themes that might emerge from the conversations. Eight of the 11 interviews were phone interviews, with each interview lasting 35 minutes on average. No interview lasted more than an hour.

The 11 follow-up interview subjects were selected based on the following criteria: (a) balance of geographic location; (b) balance of participants who had provided additional substantive information in the comments boxes, comments that warranted further clarification and exploration; and (c) balance of participants’ various roles in relation to the leadership development program (e.g., program managers, faculty/coaches, and sponsors).

The semi-structured interviews consisted of nine open-ended questions designed to elicit commentary in the following areas (see Appendix C for full list of questions and probes):

- Relationship to the program and the evaluation study (if any);
- Most valuable and least valuable aspects of the evaluation;
DISSERTATION: APPLIED META-EVALUATION

- Comments or examples related to the value of the evaluation study and the four evaluation standards.

Following the conversation with the second person to be interviewed, two additional questions were added (at the suggestion of one of the participants):

1) What are the most important elements of a successful program evaluation?

2) What role should evaluation play in a learning organization?

These additions created a broader context to open the conversation around the general role and purpose of evaluation before narrowing the focus to consider the specific Success Case Evaluation in question.

All interviews were recorded and transcribed to ensure an objective and independently verifiable record. The interviewer reminded each participant that, at any juncture, she/he could request that the recording be stopped or that a particular portion of the interview not be recorded. It is worth noting that none of the participants exercised this option or expressed any concerns about confidentiality. Participants were given the opportunity to review the transcript for their respective interviews for accuracy and the freedom to suggest edits where they felt that either the transcript was not accurate or they wanted to modify a comment. Only one participant provided any edits. All participants signed off on the transcripts, suggesting that participants were satisfied that their views had been accurately reflected. Once participants agreed to the veracity of their respective transcripts, the original recordings were deleted. Participants were offered contact information in the event they had any questions about the research or their rights as subjects in the study. Once the data for the survey and the interviews had been collected and matched, all personal identifiable information (“PII”) was removed from the survey results and transcripts and replaced with a subject code.
Archival data. In addition to the online survey and semi-structured interviews, a review of archival data relative to the program and the Success Case Evaluation was conducted to identify, corroborate, and supplement information from the survey and interviews. Documents were examined to find evidence of actions taken as a result of the original evaluation study. Given the time that had transpired since the original study, and an awareness that not all processes were likely to have been fully documented, it was expected that in some cases no archival data would be found relevant to the search criteria used.

The researcher specifically sought the following types of documents with a view that other similarly relevant documents might also be found:

- Written descriptions of the leadership program used to orient stakeholders to the program dated after the submission of the evaluation report written agreement reached with evaluation team outlining the agreed upon goals, steps and deliverables for Success Case Evaluation that was conducted;
- The Learning Impact Map (see Appendix D) that was created by the external evaluation team in partnership with the internal program manager/evaluator, which depicted the ideal impact of the training, including individual results, behaviors and capabilities needed to achieve that impact;
- The online survey that was constructed on the basis of the impact map and administered to all participants in the study;
- The results of the on-line survey;
- Interview protocol for the success case interviews;
- Notes from the evaluators from their interviews;
- Final report produced by the evaluators;
Assorted unspecified documents where changes to the program or process were made as a result of the final evaluation report.

Appendix E contains a complete list of the documents discovered and analyzed.

The archival review provided a window into how the organization prepared for and communicated the results of the Success Case Evaluation to stakeholders. Given that the review of the archival data was included as a means to verify the recollections of participants in the study, this review was conducted after the interviews were completed.

Archival data were sought from the researcher’s files and email as well as any documents that had been saved onto the internal team’s shared drive, the location where all activities related to the evaluation study would likely be archived. The four Program Evaluation Standards (accuracy, utility, feasibility and propriety) served as a guide to key areas of inquiry. Fourteen questions were constructed to complement the quantitative survey and interview data, and these questions guided the search through the relevant documents. Examples of these questions follow, with the complete list in the Appendix F:

1) To what extent was the sample of individuals selected to participate in the Success Case interviews representative of overall population of participants who had attended the training? (Propriety)

2) What was the amount of time required for the entire study (contracting through the production of the final report)? (Feasibility)

3) To what extent was the time required in line with the timing anticipated during the contracting phase? (Feasibility)

4) To what extent were the objectives of the Success Case evaluation articulated in the agreement with the external evaluators? (Propriety)
Analysis

The data collected from the online questionnaire was analyzed using basic descriptive statistics, including the mean and median scores, for each item. When the “do not know/not applicable” option was selected, the “0” score was not included in mean calculation. Additionally, two of the items (#14 and #16) were negatively phrased (See Appendix B), and so these were reverse scored in order to maintain consistency in mean calculation. Each item’s responses were organized into bar charts to review the distribution of responses against the 5-point Likert scale. Additionally, the responses to the items that fell under each of the four meta-evaluation standards were combined to create a mean score for each standard, which was compared to the means for other four categories. Given the small size of the sample, no inferential statistics were calculated.

We analyzed the data collected from the semi-structured interviews using a thematic analysis process, and archival data were consulted to corroborate perceptions where possible. Thematic analysis is a method for identifying, analyzing, interpreting and reporting patterns and themes within data (Braun and Clarke, 2006). Given the exploratory nature of this study as a review of a single evaluation versus a cross-section of evaluations, thematic analysis provided the flexibility needed to surface broader themes without the limitations of a close-ended approach derived from a more constrained data set. The process closely aligned to the phases of thematic analysis as outlined by Braun and Clarke (2006). These are:

1) Transcription of taped interviews;

2) Generation of codes and themes (using the research questions as the organizing framework);

3) Analysis;
4) Review of analysis;

5) Summary of the themes, punctuated by illustrative quotes.

The four meta-evaluation criteria research questions provided the primary organizing schema for classifying and interpreting the data. Codes were created under each of the meta-evaluation standards across the 11 interviews. These codes were consolidated into themes. While efforts were made to reflect themes based on their prevalence within the overall data set, judgments were made to include several codes or themes that seemed most relevant to the investigator, even if they occurred only once or twice.

In order to validate the author’s coding, a reviewer from the Executive Development department was engaged to determine the degree of agreement between hers and the author’s classification of comments into the respective codes. After providing the volunteer rater a short definition of each of the above-mentioned themes, the rater was asked to match 24 randomly selected statements taken from the interview transcripts to the comment category that best fit. A second space was provided to give the volunteer the opportunity, if desired, to place an alternative code if she felt that multiple codes might apply. There was overall 79% agreement.

For the archival review, the investigator began with the list of questions (see Appendix F) to be answered, first reviewing the existing hard copy documents, next the shared drive, followed by email communications in search of documentation that would provide a satisfactory answer. For each question, the researcher made a “yes” or “no” determination of whether documentation was found that would be sufficient to answer the question. If adequate documentation was located, the source of the document was cited, as well as the data that it provided to answer the question. If documentation was not found, this too was noted. It was also noted, where relevant, if the investigator had a personal recollection of the existence of a particular document, even if it
DISQUSSION: APPLIED META-EVALUATION

was ultimately not located. In those cases where there was incomplete or inadequate documentation, this was also noted. A short set of best practices relative to document hygiene will be offered as part of the Discussion section in light of insights gleaned through this document audit.
Chapter IV: Results

Meta-evaluation study results are based on three specific modes of inquiry: electronic survey, semi-structured interviews, and archival review. The meta-evaluation’s goal was to assess the worth (efficacy and impact) of the Success Case evaluation of the leadership program against the four meta-evaluation criteria: utility, accuracy, feasibility and propriety. High-level results from the electronic survey and semi-structured interviews will first be provided before considering in greater detail the results relative to each of the research questions.

Electronic Survey

Of the 21 stakeholders who agreed to participate in the study, 20 (95%) returned the signed “letter of consent” and completed the 16-item electronic survey. The author of the study also completed a survey so that there were 21 completed surveys in total. A summary of the descriptive statistics for the survey results (mean, standard deviation, and missing data) at the item level is presented in Table 1.

The overall mean scores were generally favorable, with item means ranging from 3.63 to 5.00, with the average item mean at 4.22. Mean calculations did not include those items where the “not applicable/do not know” option was selected. Specifically, there were eight items where the “not applicable/do not know” option was not utilized, while there were two items where this option had been chosen 11 times (i.e., 52% of the respondents). Generally, this option was chosen more frequently for those items that required knowledge of how the evaluation study was conducted versus those items focused primarily on the evaluation report itself.
Table 1

Meta-Evaluation Electronic Survey Results

<table>
<thead>
<tr>
<th>Evaluation Categories &amp; Survey Items</th>
<th>Mean</th>
<th>S.D.</th>
<th>N*</th>
<th>NA b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. To what extent did you feel the conclusions of the study were accurate?</td>
<td>4.57</td>
<td>.68</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>2. To what extent were the conclusions of the study clear?</td>
<td>4.38</td>
<td>.50</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td><strong>Feasibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. To what extent did you consider the evaluation to be cost effective?</td>
<td>3.80</td>
<td>1.03</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>10. To what extent did the requirements for carrying out the evaluation prove to be too time-consuming for participants in the study?</td>
<td>4.20</td>
<td>.92</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>11. To what extent did the requirements to carry out the evaluation prove to be too time-consuming in relation to the value of the finding in the final report?</td>
<td>4.09</td>
<td>1.14</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>12. To what extent did you find the delivery of the final report “timely” in the sense that the organization still had interest in the findings when the final report was distributed?</td>
<td>3.93</td>
<td>.96</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td><strong>Propriety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. To what extent did you find the questions asked in the study to be free of anything ethically inappropriate?</td>
<td>4.94</td>
<td>.24</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>14. To what extent did you encounter any bias (e.g., cultural / racial / religious / gender) in the questions asked of participants?</td>
<td>4.78</td>
<td>.73</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>15. To what extent do you believe the researchers/research team maintained the confidentiality that had been promised?</td>
<td>5.00</td>
<td>.00</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>16. To what extent were you aware of any potential conflicts of interest in the study that were not acknowledged or addressed?</td>
<td>4.88</td>
<td>.50</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td><strong>Utility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. To what extent did you find the recommendations made for program improvement to be relevant given the program and organizational context?</td>
<td>4.1</td>
<td>1.00</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>4. To what extent was the study useful to you as it related to understanding the organizational impact of the program?</td>
<td>4.0</td>
<td>.89</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>5. To what extent did you find the study’s recommendations to improve the program to be actionable (i.e., realistically implemented)?</td>
<td>3.81</td>
<td>.98</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>6. To what extent did you think that the recommendations suggested by the report, if implemented, would enhance the likelihood of participants applying the learning back on the job?</td>
<td>3.67</td>
<td>.73</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>7. To what extent do you feel the final evaluation report as it was written would be a credible document to share with different stakeholders (e.g., business partners, program sponsors, managers of participants)?</td>
<td>3.71</td>
<td>.90</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>8. To what extent are you aware of any actions taken in response to the evaluation report (e.g., changes to program content/design, communications to participants, tools, etc.)?</td>
<td>3.63</td>
<td>.96</td>
<td>15</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. 1 = Not at all; 2 = To a little extent; 3 = To some extent; 4 = To a great extent; 5 = To a very great extent

a N = the number of “scored” responses, included in the calculation of the mean and standard deviation

b NA: the number of “Not applicable/do not know responses”, which were not included in calculating the mean or standard deviation.
For example, the two items with the eleven “NA/do not know” responses had to do with whether respondents considered the evaluation to be cost effective and whether they viewed the requirements for carrying out the evaluation proved to be too-time consuming for participants in the study. In both cases, to effectively answer these questions, a respondent would have needed some knowledge around how the study was carried out, the time it required, and the costs involved. The impact of missing cases and overall item construction of the survey will be considered more fully in the Discussion section.

As noted earlier, given the exploratory nature of the study, participants were provided the opportunity to add comments for each of the items. There were 133 comments made in the open comments boxes for the 16 survey items. The question that elicited the greatest number of comments was, “To what extent did you find the conclusions of the study to be relevant given the organizational context?” This question received 14 comments, representing two-thirds of the respondents. The three items that received the fewest comments were items that were mapped to the evaluation standard propriety. Due to the small number of cases, no factor analysis was conducted; for similar reasons, no Cronbach’s alpha was calculated to measure the internal consistency of the survey.

**Semi-Structured Interviews**

Eleven semi-structured interviews were conducted, recorded, and transcribed. A thematic analysis was applied to the transcripts to generate codes and themes utilizing the four meta-evaluation standards as a starting point (accuracy, utility, propriety and feasibility) to organize the data in relation to the overarching research questions. Two additional categories emerged, which reflected comments made by participants regarding the most important elements that program evaluations should include, and the ideal role that evaluation should play for a learning
function. The resulting codes with representative quotes will be presented in relation to each of the research questions.

**Results in Relation to Research Questions**

Each of the research questions in relation to the four Program Evaluation Standards are considered below in relation to the data collected from the survey, interviews, and archival review. The extent to which the study was able to answer the two overarching research questions: (a) to what extent did the Success Case Evaluation succeed in determining the impact of the leadership development program? and (b) To what extent was the meta-evaluation useful to the organization as a means to determine the efficacy of the Success Case Method for the leadership development program? will be considered in the Discussion section.

**Research Question 1: To what extent did the Success Case Evaluation meet the evaluation standard related to propriety?**

*Survey Findings.* There were four items mapped to propriety in the quantitative survey (13, 14, 15, 16), (see Table 1). These items were the four highest mean scores on the survey, ranging from 4.78 to 5.0. In the aggregate, this category had the highest mean and lowest standard deviation ($M = 4.86$, $SD = .53$) for the responses to the four items.

*Relevant comments from the interviews.* In general terms, there were no significant concerns raised regarding the propriety of using either external evaluators or the internal group that sponsored the study as reflected in both the quantity and the nature of the comments. There were fewer comments coded under propriety in the interview transcripts (ten) relative to the other three evaluation standards feasibility (18), accuracy (41), and utility (92). Only one out of the ten propriety comments raised any concern regarding Success Case evaluation process. The subject mentioned that, as a former HR generalist, she had felt uneasy knowing that the external
researchers had conducted the phone interviews as part of the Success Case unaccompanied by an employee of the firm. Her concern was that, in the course of the interviews, a participant might raise a sensitive issue that should be addressed by someone in Human Resources (e.g., conditions of a hostile work environment, etc.). She expressed confidence that nothing like this had occurred but still felt uneasy given she did not know the evaluators well enough to be confident in their ability to either recognize or respond if they encountered such an issue.

The following verbatim comments from the interview transcripts are representative of the overall comments related to Propriety. Each comment comes from a different interview respondent as a single, unique expression, unless otherwise noted. If a number in parenthesis follows the statement, the number represents similar sentiments expressed somewhat differently.

- I had no concerns whatsoever (6).
- I had no concerns [ethical in nature], as participants had the opportunity to self-select out if they felt uncomfortable by the process or the questions.
- I had no [ethical] concerns, as the approach and questions were unbiased.
- I trusted the [internal] team and the fact that we had used an external vendor.
- Didn’t have line of sight into how the study was conducted, but trusted the internal team to avoid any ethical issues or conflicts of interest.
- My only concern was that researchers might surface ethical issues in the course of interviews and we might not be aware of them since none of us was on the phone with them during the interviews.

Archival data. No documents or correspondence were found during the archival review of communications from the internal evaluation team indicating any concerns relative to the ethics or propriety of the items or the way in which the study was conducted. The working
agreement between the organization and the evaluators did not specifically outline any ethical considerations, although the non-disclosure agreement ("NDA") promised the mutual confidentiality of all information that was shared between the evaluators and the organization.

Research Question #2: To what extent did the Success Case Evaluation meet the evaluation standard for accuracy?

Survey findings. There were two items mapped to accuracy in the quantitative survey (1 & 2), see Table 1 - Evaluation Survey Results (p. 54). These items represented the fifth and sixth highest mean scores on the survey, with means of 4.57 and 4.38. As a category, accuracy had the second highest mean and second lowest standard deviation ($M = 4.48, SD = 0.59$).

Relevant comments from the interviews. The category accuracy generated the second highest number of comments in the interview transcripts (41 indicating that the main conclusions of the evaluation were credible, trustworthy, and based on good quality data. Additionally, there were a number of comments from interviewees suggesting elements they would have changed or found missing in terms of the Success Case evaluation report, which will be outlined in the representative comments below. (e.g., more longitudinal data, the ability to see all comments from participants, etc.).

- I agree with the main conclusion of the report – the program could have had more impact (4).
- Conclusions were trustworthy and justifiable by the data (3).
- The study was more rigorous than others I’ve seen conducted here (3).
- Found conclusions to be logically sound and consistent with my own experience (2).
- I found the results to be trustworthy because the data reported a not overly positive picture.
DISSERTATION: APPLIED META-EVALUATION

- Provided a balanced view of the course, but was missing the motivational and positive attitudinal elements the course had on participants.

- The study was trustworthy but more a reflection of how we do leadership development than the specifics of the course.

No one disagreed with the broad conclusions of the study, but over half of the interviewees expressed some desire for the study to have gone further in detailing its conclusions. Amongst the comments around the limitations of the study, one individual felt that the original survey could have been stronger, as there were a number of “double-barreled” items that lacked precision. Another interviewee felt that having more participants and longitudinal data would have enhanced credibility.

Archival data. The first document reviewed in relation to accuracy was the Learning Impact Map (See Appendix C), which the internal corporate team and the external evaluators collaboratively created in order to inform both the survey-item construction and interviews that would be used for the Success Case. The second document to be reviewed was the survey itself. (See Appendix G ?). The survey had an acceptable response rate of 66% (200 of 299 completed it) and the entire population was given opportunities to participate.

These archival data presented several challenges. The external evaluators’ notes from the interviews conducted were not obtained for this study, and may have provided additional insight into the decision-making process that was used to determine which quotes and Success Cases were ultimately included in the final Success Case Report. Also, the external evaluators had some difficulty scheduling the interviews, but it is not clear how this impacted the selection of

---

11 A double-barreled question is one which conflates more than one issue, but allows for only one answer, creating possible confusion for the respondent and for the item’s interpretation. It should be noted that the investigator upon reviewing the survey in Appendix G did not find any such items, although there were double-barreled response options.
Success Cases nor is it certain whether this reflects any reluctance on the part of the organization’s employees to participate, or whether this was simply a reflection of limited available time by external evaluators’ and/or participants.

**Research Question #3: To what extent was the Success Case Evaluation a feasible approach for the organization?**

**Survey findings.** There were four items mapped to feasibility in the quantitative survey (items 9, 10, 11 and 12), see the Table 1. As a category, feasibility was ranked third out of the four evaluation standards in terms of category mean and had the highest standard deviation of any of the categories ($M = 4.04$, $SD = .98$). The four items associated with the category had a mean range from 3.8 to 4.2. While these ranged lower than the overall mean for all items ($M=4.21$), objectively these still represent largely affirmative responses to the questions.

The higher standard deviation for the category is likely a reflection of two of the items (9 & 11), where 11 of 21 or 42% of the respondents utilized the “do not know/not applicable response,” which meant that there were fewer respondents who indicated an answer choice on the response scale that was calculated into the mean and standard deviation. Item 9 specifically had to do with the “cost-effectiveness” of the study, and only a very small subset of stakeholders (3) in the meta-evaluation would have had first-hand knowledge to answer this question.

Similarly, item 11 was related to whether the requirements of the study were too time-consuming in relation to the overall value of the reports’ findings. For those not acquainted with whatever efforts were required, answering this question would have elicited either speculation or required specific second-hand knowledge. The utility of including items that presupposed knowledge beyond the Success Case evaluation report will be addressed in the Discussion section.
**Relevant comments from the interviews.** The category feasibility elicited the second fewest number of comments in the interview process (18) relative to the other categories. Many stakeholders prefaced comments with an acknowledgement that they had limited line-of-sight into the actual work involved to conduct the Success Case evaluation. Others reflected on the conditions in place for the organization at the time of the original evaluation and speculated pessimistically about whether the study would be replicable in the organization’s current environment. Others noted that the effort required to conduct a study of this nature is often overestimated and thus serves as a barrier to even making the attempt. The general consensus amongst interviewees was that the requirements to participate in and/or carry out the study were reasonable and not excessive, given the investment made in participants and the visibility of the program within the organization.

Interviewees interpreted the participation rate in the Success Case survey itself as a positive indicator of feasibility. The following are representative quotes:

- The amount of organizational effort to conduct the study seemed reasonable, but I may be wrong (3).
- The time it took from the start of the study to receiving the final report did take time, but was within a window that I’d consider reasonable.
- Not sure the organization would have the same appetite for this kind of study now given all of the surveys that we ask people to complete.
- The requirements and expectation of participation are reasonable given the investment of the company in the individuals.

**Archival data.** The original timelines outlined in the work order for the study were examined in the archival review. The review revealed that the evaluation took two months longer
than the originally anticipated three months. The person who was internally responsible for the study suggested that the primary cause was the challenge of securing the follow-up Success Case interviews between alumni from the program and the external evaluators. As noted above, the Success Case response rate to the electronic survey seems to indicate that it was reasonable to expect participation in this first phase, while the small number of interviews completed in follow-up, seems to reflect the more difficult of the two steps (survey plus interview) required to implement this method for program evaluation purposes.

**Research Question #4:** To what extent did the Success Case Evaluation meet the evaluation standard for utility in the context of the organization in which it was conducted?

**Survey findings.** Six items were mapped to utility (Items 3-8), see Table 1, as utility represented the greatest interest to the study. As a category, utility had the lowest category mean and the second highest standard deviation of any of the categories ($M = 3.98$, $SD = .88$). Four of the five least favorably scored items were associated with utility, ranging from 3.81 to 3.63. This latter item had to do with the awareness respondents may have had regarding any subsequent actions taken as a result of the report’s findings. It should be noted however be that while 3.63 is the lowest scoring item on the survey, objectively it falls within “to some extent” and “to a great extent” on the 5-point Likert scale.

**Relevant comments from the interviews.** The category utility generated 91 comments, more than twice the number generated by the category accuracy, which generated 42 comments. The category utility was classified into four subcategories in order to better understand the different aspects presented. Each subcategory is followed by illustrative quotes:

*Limitations of the study and changes that would have enhanced the utility of study or the final evaluation report (36).*
The utility of the study was limited because the study provided limited new insights (we could have guessed what the conclusions would have been and perhaps these could be said of all our programs not just this one specifically).

The utility of the study would have been higher with more specific recommendations and sharing of best practices.

The utility would have been higher if they (the external evaluators) had given us more creative solutions…they gave us obvious answers that we know haven't worked. I expected more since they were external.

The utility of the study is contingent upon action and this depends a great deal on the organizational context at a given moment in time (vs. just the report itself).

Greater depth and color about the success cases and where impact was being felt.  

*The study’s greatest value (21).*

It helped confirm things we knew but gave us data to help tell and back up the story.

It gave us the general sense of whether or not we were reaching the objectives set out by the program - I thought that was valuable data.

What was most valuable was that the study went beyond Level 1 to look at impact.

Validation of the need to focus on the system and not just the content.

It confirmed that we are leaving a lot of value on the table.  

*Awareness of actions that were taken as a result of the Success Case evaluation (19).*

Included the study’s results in the program itself.

I recall there were some actions taken associated with engaging the manager or being clearer about nomination process with participants.
The study’s results were shared more broadly (not sure of overall impact of this beyond awareness).

I assume actions were taken in response to the report - e.g., kept table coaches as a result of the study.

I was aware of ongoing changes that were aligned with the study's conclusions, but not necessarily driven by the conclusions.

We made adjustments to pre-program communications to managers.

*Contributions the study made to the organization and/or the Learning function (10).*

Demonstrated effort to show ROI.

The study signaled a more professional, business-minded L&D function.

The focus on measurement reflects positively on the Learning department.

There was value in thinking about how to increase the participation of participants' managers as a means to create more sustainable skill development.

Raised visibility of program in organization - important as a new program.

All but five comments were coded under the above four subcategories. The additional five did not fit the above categories and were coded as miscellaneous.

*Archival data.* Documents were sought to validate any changes that had been made to the program design or program processes (e.g., communications before and after the program) as a result of the evaluation report’s recommendations. Evidence was found that the study’s findings were included in the facilitator guide and program slides during the program and also in webinars held pre-program for the managers of participants. Evidence was found that the Evaluation Report had been disseminated and discussed with all of the program’s facilitators and
Due to the exploratory nature of the study and the alteration of the interview protocol, two additional but unanticipated themes emerged from the interviews. While these themes do not directly address the research questions, they do shed light on the thinking of the stakeholders who were interviewed relative to the purpose and desired outcomes of evaluation efforts in corporate settings. These themes were coded as follows:

**Defining evaluation success – key elements.** Interviewees, in addition to commenting directly on the Success Case evaluation as part of the meta-evaluation study, shared views on what they would consider to be a valuable evaluation. There are strong connections between these additional comments and those relative to the Success Case evaluation and how its utility might have been increased. One element emerges from these comments: Successful evaluations provide quantifiable evidence of behavior change, business impact, an objective view of the extent to which these can be attributed to the program, and guidance around how to address barriers and enhance the program’s ability to facilitate change and impact. The following are specific comments from the interviews:

- The most important element is to be able to measure that program attendance led to behavioral change in the workplace.
- Can we measure if leaders who attend are more capable and prepared?
- Measurement of behavior change from the perspective of key stakeholders.
- Should inform how to improve outcomes, measure the effectiveness to drive those, and improve the quality and effectiveness of the program.
- Should measure intent, actual impact and the gap between the two.
Should measure the degree of learning and application.

Should correlate (program objectives) against real data versus tracking stated intent.

Sometimes we overlook the emotive element, that is, how the program made the participant feel about themselves and the company; that is, are they more engaged and committed as a result of having attended the program?

**Theme - the ideal role for evaluation within a learning organization.** Respondents also made several comments suggesting that they believe learning organizations must continue to evolve and invest in evaluation efforts to more clearly understand impact, meet growing stakeholder expectations, and be perceived as credible partners and professionals.

- Evaluation should play a more prominent role in learning functions - we invest disproportionately in planning and execution.
- Focus on metrics and evaluation is important professional obligation of the training function.
- Evaluations that provide quantitative data back to key stakeholders, like program sponsors, can help shift the onus of responsibility for results back to the business where it belongs (vs. in HR or Learning).
- Our focus in evaluation is often weighted far too heavily on program experience and not program effectiveness.
- We have a responsibility to the business to speak to them in their language and communicate results in quantifiable terms.
- Stakeholders are generally demanding more in terms of demonstration of value (ROI, increased productivity, improved team functioning or relationships).
Having now reviewed the results coming from the three sources of data in relation to the research questions related to the four Program Evaluation Standards, we return to answering the two overarching questions posed at the outset of the study:

1. To what extent did the Success Case evaluation succeed in determining the impact of the leadership development program?

2. To what extent was the meta-evaluation useful to the organization as a means to determine the efficacy of the Success Case Method for the leadership development program?
Chapter V

Discussion

This study applied meta-evaluation criteria and standards to assess the worth (efficacy and impact) of an already completed Success Case Evaluation of a leadership program in a global bank. The primary purpose of the study was to determine the extent to which the Success Case Evaluation met the four meta-evaluation criteria: accuracy, feasibility, propriety and utility. Of these four, the study was most concerned with determining the feasibility and utility of the Success Case Method as applied to the leadership program. Second, the study also attempted to gain a preliminary understanding of the role meta-evaluation itself may play as an ongoing discipline to inform the organization’s evolving metrics and evaluation strategy. This secondary objective will be considered in final conclusions as an extension of the second overarching research question, “To what extent was the meta-evaluation useful to the organization as a means to determine the efficacy of the Success Case Method for the leadership development program?”

In this discussion section, the results of the study will first be considered as a whole, assessing the extent to which the primary research questions were answered. Consideration will be then given to understanding how these findings relate to extant academic and professional evaluation research, implications for the field of training, limitations of the study, and some implications for future research. In addition, a section will be included to review the author’s role as participant-researcher in the study.

Summary of Findings

Main Findings. The first overarching research question evaluates the extent to which the Success Case evaluation succeeded in helping the organization understand the impact of the
DISSERTATION: APPLIED META-EVALUATION

leadership development program. Overall, the results from the quantitative survey and semi-structured interviews largely affirm that participants viewed the evaluation as having provided a clear, high-level snapshot of the program’s impact and the key factors that limited impact. The Success Case’s most important contribution was to provide concrete data to ascertain the extent to which program participants were applying what they had learned in the training to their respective jobs. Item mean scores on the quantitative survey were positive and ranged from 3.63 to 5.00 on a 5-point Likert scale with an overall mean of 4.22. There was universal agreement with the SCM’s main conclusions. Similarly, a preponderance of comments affirmed the evaluation’s findings were accurate, trustworthy, free of bias, and generally aligned with their own views and experiences regarding the program’s impact. The conclusions, which were most commonly cited from the report, were:

- The company was “leaving money on the table;”
- Managers need to be more involved;
- Only a small percentage of participants were applying the learning to obtain business results.

The conclusion that the company was “leaving money on the table” was a central theme highlighted in the evaluation report and warrants some explanation. By this, the evaluators signaled that the company was not reaping the full benefit (impact) that it might expect based on participants’ overwhelmingly positive response to the program and the level of investment the company had made in the experience.

While the data from the surveys and interviews supported the conclusion that the Success Case evaluation was viewed as accurate, proper (possessed a high degree of propriety), and feasible, there was no strong consensus around its utility. While there was no suggestion that the
Success Case evaluation was devoid of utility, many responders considered the potential utility to be limited.

Despite mixed views on the overall utility of the study, participants indicated that the Success Case evaluation had made important contributions to the learning function. They noted that the Success Case evaluation represented a more ambitious and comprehensive attempt at program evaluation than the organization had been previously undertaken, going beyond the standard reports of attendance and participant reaction to the training (Kirkpatrick’s Level 1). Several participants in the interviews expressed that the initiative positively signaled increased professionalism of the Learning organization while simultaneously underscoring the importance of the Leadership program itself.

**Evaluation standard findings: propriety.** Stakeholders largely agreed that the Success Case study had been conducted in an ethical manner and was absent of any significant biases or conflicts of interest. The stakeholders’ perceptions of the credibility of the study was generally enhanced by the use of external evaluators, suggesting the evaluators had brought an impartial objectivity and subject matter expertise to the evaluation. In addition, participants expressed that they were confident that the internal team would have addressed any ethical issues had any arisen.

As mentioned previously, one individual did raise a specific ethical concern. Her concern was that there had been no one from the organization present during the interviews with the external evaluators. She felt that participants might, in the course of the interviews, raise issues that have ethical components and that those issues might either go unrecognized by the evaluators or be managed ineffectively. Her concern was valid and highlights the desirability of having internal personnel participate in the interviews with stakeholders. Two possible ways to
address this concern are: 1) the internal team (which included the investigator) could have provided the evaluators with an approved script that they could communicate to participants relative to how information would be used following the interviews and indicating what actions would be taken if any ethical issues were uncovered or raised in the interview process; or, 2) the internal team could have accompanied the external evaluators in their phone interviews.

These possible changes would have addressed the concern, but would have had some impact on feasibility, given the greater time and internal resources required to complete the evaluation. While it is clear that the evaluators did broach the subject of confidentiality and its limits with participants, no documents were identified through the archival review indicating the nature of what was communicated. Nonetheless, the concern raised signals the importance of anticipating, documenting, and clarifying expectations and processes to be followed should an ethical issue surface in the course of the interviews.

**Evaluation standard findings: accuracy.** Accuracy relates to the standards which are meant to ensure that the evaluation will reveal and communicate technically defensible information, lead to justifiable conclusions and deliver impartial findings. In this respect, stakeholders raised no serious concerns in regards to the accuracy of the conclusions of the Success Case evaluation or the methodology used to obtain them. As was the case with propriety, a level of expertise and thoroughness was assumed by stakeholders in relation to the external evaluators. Many had some familiarity with the Success Case Method and were comfortable with the approach and the knowledge that the method’s creator had sanctioned the evaluators to apply the method. Perhaps more influential, however, was the fact that the findings of the study were consistent with their own views, namely that “money was being left on the table” and that increased manager involvement would enhance the likelihood that participants
applied what they had learned in the training. While this phrase was not actually used in the evaluation report, it came up three times during the interviews, which seems to reflect how that finding from the evaluation had been internalized. The actual statement in the evaluation report was “the company is leaving considerable impact on the table from the leadership program.” Both the survey data and the interviews reflected a strong sense that the findings and conclusions of the Success Case evaluation were viewed as accurate and justifiable, and therefore, trustworthy.

Five participants, however, raised particularly thoughtful questions during their interviews. One noted that the findings of the evaluation could be considered “accurate” assuming the questions that were asked were the right ones. In other words, based on the program objectives, the participant considered the evaluation to be accurate. For this participant; however, the purpose of an evaluation ought to go beyond trying to understand the extent to which knowledge was acquired and applied. More specifically, this individual held that effective training should also aim to awaken a desire for ongoing learning in participants, and so an evaluation ought to capture changes in mindset and orientation toward learning in participants. We will return to the question of “mindset” as a component of interest for evaluation in the discussion on future direction in the research.

Another participant, who expressed a high degree of confidence in the accuracy of the Success Case evaluation, did suggest that the evaluation would have even greater accuracy if the Success Cases had been validated by views of other stakeholders. This person suggested a process that more closely resembled a 360-degree evaluation that would take into account the manager and direct reports of the person who attended the course versus a self-report from the participant.
A third participant in the study raised the question of confirmation bias. Because the conclusions of the Success Case evaluation were aligned with the predominant view held amongst members of the Learning Function (that a general lack of manager involvement was limiting the impact of the program). This individual wondered whether the organization had been too eager to accept this conclusion without pressing further to understand if there were other important drivers or impediments to greater learning transfer that the evaluation had missed.

A fourth person raised questions around the construction of the survey items used in the Success Case evaluation. This participant noted that some of the electronic survey items had been double-barreled (i.e., essentially asking two questions in a single item), which would have obscured the clarity of responses, and by extension, the findings themselves. This participant also suggested the study’s conclusions would have been more credible had there been longitudinal data regarding the performance of the participants over time, recognizing that this was not part of the evaluation’s original scope. Lastly, given the evaluators had never operated within the organization, this individual expressed that having access to the evaluators’ interview notes (vs. only selected quotes) would have further increased her confidence in the evaluators’ conclusions.

This last theme was picked up by a fifth participant in the meta-evaluation interviews who shared that while he trusted the data and the representative nature of the quotes included in the final report, he would have preferred to see an appendix with an exhaustive list of all quotes, as he could then draw his own conclusions based strictly on the data.

**Evaluation standard findings: feasibility.** Feasibility refers to assurance that the evaluation is practical, viable and cost effective. The methodology used and processes implemented must take into account the organizational context and be carried out in ways that do
not disrupt organizational routines or be viewed as overly intrusive. This criterion is pragmatic in that it seeks to assess whether the benefits of an evaluation approach warrant the effort and resources required to obtain them. Factors impacting feasibility include: 1) the amount of time to conduct the study; 2) actual costs, which includes both internal L&D resource expenditures as well as external fees paid; 3) the degree of organizational sponsorship; 4) the intrusiveness of the evaluation; 5) company performance; 6) the timing of the evaluation; 7) the composition of the stakeholders involved; and, 8) available resources. Given the purpose of this study was aimed more at the evaluation of a method versus broader questions of feasibility, the meta-evaluation sought to answer questions related to the evaluation approach itself (the Success Case Method) versus the broader organizational factors mentioned above (e.g., current economic performance of the organization, degree of sponsorship, etc.).

The feasibility-focused questions were designed to ascertain perceptions of key stakeholders relative to the financial cost, amount of time the evaluation study took, and involvement required from stakeholders within the organization (i.e., in this case, the internal evaluation team and participants in the Leadership program). By including feasibility as one of the foci of this research, the goal was to gain an appreciation of key stakeholders’ perceptions of the value of the study relative to the Success Case’s feasibility to carry out, which would also factor into the likelihood that the organization would choose to deploy the method again in the future.

Participants generally agreed that the Success Case evaluation had been feasible to conduct (M=4.05, SD =.99). However, it is clear from the number of times the “do not know/not applicable” option was selected for several items (38 times in total), high standard deviations for this evaluation standard, and the scarcity of comments in the interviews, that participants were
not confident in making strong statements in relation to feasibility. In the interviews, for example, four participants prefaced comments by saying either, “To the best of my recollection,” or “I’m not sure, but I think,” etc.

The relative lack of specific information from participants related to feasibility leads to the following observations. First, since all but two of the participants in the study had not been part of the internal team that had assisted with the evaluation, participants were not in a position to feel confident in their responses to questions around feasibility. They did not feel they had a clear line of sight into how the study was conducted. Second, given the amount of time between the Success Case evaluation and this study, it is unreasonable to expect that those individuals who were not directly involved in the evaluation itself would recall elements relative to feasibility. The decision to include feasibility as a standard, however, was an important decision in the research design, as feasibility is, by definition, a key variable in the decision to sponsor an evaluation or deploy a particular method. By including this standard, the researcher hoped to uncover whether any of the key stakeholders would have either assumed it was a very costly study and whether they viewed it as labor- or stakeholder intensive. It was also meant to uncover whether there were any perceptions of delays in relation to the time required for the study to be carried out.

While knowledge of key stakeholders’ perceptions is instructive with the aim of setting effective expectations for future studies (e.g., if the “brand” of an evaluation approach is “time-consuming,” “expensive,” or “intrusive”), the inclusion of feasibility in this study could have been more efficiently ascertained by consulting the internal team and related documentation, particularly given the time lag between the SCM and the meta-evaluation. Taking all of this into account, what is clear is that there were no significant concerns from the stakeholders relative to
the carrying out of the Success Case evaluation, the caveat being that these reflections were
given in hindsight and were thus subject to memory.

**Evaluation standards findings: utility.** Utility refers to the usefulness or ability of the
evaluation to serve the information needs of the intended users. Based on the investigator’s
personal knowledge of the organization, it was anticipated that utility would be the most
important variable to stakeholders. The number of open-ended comments that were classified
under this standard in the qualitative interviews (92), more than twice the number of comments
associated with the nearest category (42) supports this assertion. Also, given the number of
comments, utility was the meta-evaluation standard for which there was the greatest variability
of perceptions in both the survey and qualitative interviews.

Based on the data results, participants generally agreed there was utility to the study but
that it was limited. This is based on almost universal agreement with the study’s primary
conclusion (“considerable impact being left on the table”), the diagnosis of the cause
(insufficient direct manager support), and the concrete data the study provided to support its
findings. Four of the participants noted that the Success Case evaluation was the most
comprehensive and systematic evaluation that they had participated in within the organization.
Three others made reference to the fact that the Success Case went beyond “Kirkpatrick Level 1”
to focus on impact, which they viewed as positive.

There were several changes that participants in the interviews remembered as having
happened as a result of the Success Case evaluation. Some of the implemented changes noted
were:

- Communication of the results of the study to the faculty and stakeholders of the program;
Inclusion of human resource generalists in the pre-program calls for managers to increase awareness of the program;

- Insertion of the results of the study citing factors, which were associated with greater impact (manager and coach engagement) as part of the pre-program calls with managers, the facilitator guide, and into the program materials.

One participant who was the Head of Leadership and Management Development noted two additional impacts of the study. First, the coaches began to take on more ownership and push harder to continue engagements with participants, as they saw evidence in the evaluation report that continuing with the coaching had a positive impact. The second is that the Success Case evaluation established a benchmark and became a source of review and stimulus for improvements to other core leadership programs and evaluations.

The perception of utility was limited by a number of factors. First, participants indicated that while both the diagnosis and conclusion were consistent with their own perceptions, some felt that these conclusions were not unique to this particular leadership program. Rather, the same conclusions would be equally applicable to other leadership programs offered within the company. Second, while there was agreement that greater management involvement was an important lever to increase the impact of the program, the evaluation’s recommendations for improvement were too generic and not contextually specific. One person commented, “We’ve known that this [lacking manager involvement] has been a challenge for years, but whatever we’ve done hasn’t addressed this, so I was hoping we would have more by way of recommendations.” Another interviewee commented that the “success cases” weren’t clear or specific enough to be able to identify ways to unlock greater application of learning from the program. Another participant in the interviews suggested that the recommendations could have
been tailored to different stakeholder groups (participants, participants’ managers, program managers, sponsors, etc.) to increase the likelihood that action would occur.

One assumption connected to the meta-evaluation participants’ overall evaluation of utility had to do with the number changes (to the program design or otherwise) that had been made as a direct result of the evaluation study. As one participant put it, the utility of the evaluation is/should be “contingent upon action and this depends a great deal on the organizational context at a given moment in time.” For this participant, and others, the true test of the value of an evaluation is whether it leads to meaningful changes to program design or structure. While this viewpoint has some validity, it is important to note that the ability (or inability) of an organization to act on recommendations from an evaluation is dependent on many dynamic factors in the organizational context. Therefore, to judge the utility of an evaluation solely based on any subsequent actions inspired by the evaluation would place unfair burden on the evaluation. This is not to suggest that an evaluation that produces recommendations which are misaligned or poorly calibrated with the organizational context should be excused, but simply that the utility of the evaluation should not be held hostage to whether or not the recommendations were ultimately implemented.

First, the comments in the preceding paragraph help us situate the Success Case evaluation as primarily being formative in nature. It is clear that the stakeholders interviewed were not only interested in assessing the impact of the leadership program, but that the overarching goal was to understand what elements of the program design and company culture could be improved to ensure that not only this program’s impact was maximized in the future, but that other programs might have greater impact as well.
It is also worth recognizing that organizational context plays an important role in determining the ability of an organization to implement any change, including those related to implementing the recommendations of a program evaluation. Specifically, what are those elements in the system that might adversely impact an organization’s ability to act upon the results of an evaluation? Or put more positively, what are those organizational levers that might be utilized to enhance the likelihood that change occurs?

**Burke-Litwin (1992) Model of Organizational Change**

The Burke-Litwin model of performance and organizational change provides a helpful framework for understanding these questions. This model was first proposed by Warner Burke and George Litwin (Burke and Litwin, 1992) as a means to wed theory with practice in a model that is intended to not only be descriptive but also diagnostic. Burke and Litwin proposed an open-systems understanding as means to understanding the dynamic interaction between 12 key dimensions that impact performance and organizational change. These 12 organizational dimensions are hierarchical and take into account both internal and external variables. These include: the external environment, mission and strategy, leadership, organizational culture, structure, management practices, systems, work unit climate, task and individual skills, individual needs and values, motivation, and individual and organizational performance. The most dominant factor of these is the external environment, which exerts pressures which impact changes to the organization’s mission, culture, leadership, organizational culture, and structure, etc., through the 12 dimensions. As can be seen in Figure 1 below, those dimensions on the upper portion of the diagram exert greater force as factors affecting change than those in the bottom half.
Burke and Litwin also make a distinction between transformational and transactional change. Transformational change happens as a response to the external environment and directly affects mission and strategy, leadership and organizational culture. These, in turn, affect the transactional dimensions: structure, management practices, systems and climate. Together, transformational and transactional factors affect individuals’ motivation, which in turn, has an impact on individual, team and organizational performance.

*Figure 1*: Burke-Litwin Model (1992) of Organizational Performance and Change
How then can this model help explain some of the challenges of the organization to implement the recommendations from the Success Case evaluation? In the first place, one of the positive elements of the Success Case Evaluation was that it sought to address the organizational system by focusing not only on the program participants and their motivation to apply learning toward business results, but also on the participants’ managers. Returning to the Learning and Development Roundtable Study (2008) cited earlier, researchers found that manager feedback and communication with participants after a program exerted a 17% impact on motivation to apply learning back on the job. As noted, however, the focus on managers is a necessary but not sufficient condition for change. The model situates management practices, motivation, individual needs and values, and work climate as transactional factors. As such, they are largely (although not entirely) at the mercy of the transformational factors.

In contrast, the organization in which the leadership program was delivered is in financial services, an industry that at the time of the program pilot and Success Case evaluation (2009-2010) was struggling through the impact of the Great Recession of 2008 and in the midst of transformational change. As a result of the crisis, external regulatory and governmental bodies began to exert greater influence over the organization’s direction with dramatic changes that were meant to prevent future financial crises. As would be anticipated by the model, the organization responded by making important changes to its leadership (new CEO, senior leadership team, new members to the Board of Directors), strategy, which involved exiting many businesses, organizational structure, systems, policies and practices. While organizational culture is, according to the framework, a transformational dimension, with all of the changes occurring, a shared consensus around the organization’s culture was hard to identify amidst the change. It should be noted that not all of the changes mentioned above would be counter-productive to
individuals and their managers feeling more accountable to seek business impact after attending
a training program. The senior leadership team sent important messages to the organization
about the importance of learning and the continued investment in the workforce throughout the
crisis, but this was not sufficient to overcome the organizational inertia (i.e., there generally had
not been much manager involvement or sense of individual obligation to encourage program
participants to apply learning from leadership programs back on the job prior to the Success Case
evaluation or the crisis) or the general sense of insecurity that prevailed throughout the
organization as the result of workforce reductions and the additional responsibilities assumed by
those who remained.

It is not surprising then, when taking into the account the preponderance of changes
taking place at the transformational level, that efforts to incentivize and catalyze new behaviors
aligned with the program’s objectives (e.g., more effective and frequent coaching of direct
reports) at the individual and manager levels were less impactful than hoped. Nonetheless,
changes to the leadership program as a result of the Success Case evaluation were made in line
with recommendations, although the hoped for impact of greater manager involvement and
subsequent business impact has been unclear at best.

What ways would the Burke-Litwin model suggest as ways to incentivize individuals and
their managers to seek greater application and business impact from training in the future? First,
the model would suggest that those responsible for managing the training function would
consider the external environment and the transformational dimensions in addition to the
transactional. For example, it would be helpful to pay careful attention to any cues of challenges
in the external environment where effective leadership behaviors might enhance the
organization’s ability to meet its challenges. Second, where there is an environment of great
uncertainty and finite resources, even greater efforts must be made to influence visible senior
developer support and involvement in the programs themselves and articulate the expectation that
participants make application from the program back on the job. While the training function
may run the programs, accountability for results should reside with the businesses. Third,
expectations need to be made more explicit relative to specific responsibilities that managers of
participants have in relation to their role as sponsors of their direct reports’ development. This
would include having structured conversations with their direct reports to both reinforce and
identify specific opportunities to apply what they’ve learned on the job. This could be tracked
through a follow-up survey at 3 or 6 months after attendance at the program, and should factor
into the annual performance review for both participants and their sponsoring manager. Lastly,
the training function must continue to do the work at the transactional level with managers and
individuals to identify and showcase positive examples of how learning from programs has been
used to have business impact.

**Overall Value of the Meta-Evaluation for the Organization**

The second overarching research question asks, to what extent is the meta-evaluation
valuable to the organization as a means to determine the efficacy of the Success Case Method as
applied to a leadership development program? Overall, the meta-evaluation was an effective
means of understanding the value of the Success Case evaluation of the leadership development
program. It also served as a catalyst for rich reflection around the role of evaluation more
broadly in the organization. It was, however, limited its ability to critique the Success Case
Method itself, as participants in the study did have the opportunity to review raw data, but
instead reviewed the Success Case Evaluation Report that presented data at a high level.
Participants in the meta-evaluation were not given a formal explanation of the Success Case
Method and the assumptions behind the approach, and as a result most of the comments from the interviews focused on the evaluation itself and not the method.

The meta-evaluation did surface many useful views in relation to the Success Case evaluation. These views, if incorporated, would inform and refine the focus of both future evaluations and of meta-evaluations. For example, one interviewee mentioned that an important variable often missing from evaluation is the measure of the way that participants feel about the company as a result of the company’s direct investment in them. Another interviewee mentioned that an important outcome of evaluation goes beyond job-specific knowledge and skills and into an overall stronger commitment and orientation toward personal development. This particular comment underscores the important point made by Cherniss and Goleman (1998) that effective leadership development needs to address social and emotional learning, and the added complexity which requires that the learner be ready and motivated to the change. The comment also seems aligned to Stanford Professor Carol Dweck’s work on the concept of mindset.

Dweck (2006), in her book *Mindset: The New Psychology of Success*, highlights two basic mindsets that she has identified in her research. The fixed mindset is based on the belief that one is born with a static amount of intelligence or talent. The growth mindset, in contrast, is based on the belief “that your basic qualities are things that you can cultivate through your efforts.” (p.6). Her research has found that those individuals possessing a more of a fixed mindset are risk averse, tending to experiment less and take less risks, instead seeking to protect themselves against failure. In contrast, those with more of a growth mindset tend to view success as stemming from hard work, learning, and persistence. Put another way, the fixed mindset sees situations in binary terms around success and failure, whereas the growth mindset sees situations as a spectrum of opportunities to learn, grow, and progress toward mastery. If
this is true, it makes sense that one of the objectives for training would be to help cultivate a
growth mindset, which will have a longer-term impact on learning and performance than the
specific knowledge, skills or abilities covered in a given program. In particular, the notion of
learning through experimentation and a reframing of failure could be built into the program
design and post-program application. Program design could include, as pre-work, concepts that
focus on mindset, sharing how growth occurs and how program participants can better anticipate
and overcome challenges to growth. This could also include a focus on how to prevent relapse
into old behaviors and habits after the program. If these form part of a program’s overall training
objectives, it would be worth exploring how to measure the desired change in mindset (e.g., pre-
test/post-test).

Similarly, the meta-evaluation and archival review surfaced opportunities to strengthen the Success Case evaluation without making changes to methodology. For example:

1) In the future, if using external evaluators, a more formal review of roles and
responsibilities and the interaction model between the external evaluators and the
organization should be clarified.

2) While confidentiality was communicated with participants, there could also be a more
formal discussion between the internal and external evaluators to determine how any
potential ethical issues might be managed if they were to surface.

3) The meta-evaluation brought to light that it would have been valuable to have a broader
group of stakeholders involved in shaping the objectives of the evaluation and also in the
dissemination and implementation of recommendations. Specifically, broader input into
the creation of the Learning Impact Map (see Appendix C) would have potentially led to
a more refined set of survey items.
4) With the knowledge of the importance of manager involvement coming from the Success Case evaluation, future evaluations could spend more time truly building out the details of success cases versus those where individuals had not taken actions to note any practices or behaviors that would be helpful to foster greater impact.

5) In the future, involving the managers of participants in the survey to validate impact would serve not only as a means to overcome self-reported data from participants, but would also serve to remind managers of important actions they need to take before and after a direct report attends a program.

6) While it was clear that there was an action plan in place on behalf of the internal evaluation team to both communicate and implement the Success Case findings and recommendations, this plan did not include collecting more data in the future to measure the effect, if any, these efforts had made to program impact.

7) The Success Case evaluation report could be enhanced by the inclusion of an executive summary as well as a full list of quotations coming out of the interviews. This was feedback that emerged from the interviews and makes a great deal of sense. Where possible, individuals identified as “success cases” could provide testimonials, participate in pre-program calls with participants and their managers or even film short clips detailing what they did that helped them to apply what they learned and what impact it had on the business, their teams and themselves personally.

8) Future evaluations should also consider whether tracking the extent to which an individual felt positively about the company as a result of having had the opportunity for the experience or the extent to which she had developed a commitment to ongoing development, were two suggestions from the meta-evaluation interviews worth exploring.
Perhaps most importantly, the meta-evaluation provided insights that could be applied broadly to the practice of evaluation within the organization. These insights ranged from tactical elements, like those mentioned in the previous paragraph, to the more strategic questions around the very purpose of evaluation. In addition, the interviews for the meta-evaluation highlighted the growing desire to further connect and integrate evaluation results with the organization’s own performance metrics, such as performance ratings, attrition, retention, promotions, mobility, and the company’s voice-of-the-employee surveys. Not only did the stakeholders suggest that this integration should happen, but that individuals should be tracked over time, so that application and impact could be viewed over the life-cycle of employees. While the meta-evaluation itself did not provide specific answers, it has helped raise questions that could aid the organization in the development of its metrics and evaluation practices.

While there were no specific questions in the interviews or surveys related to the value of the meta-evaluation approach as applied in this case, the investigator did ask five of the respondents a question around the potential value of meta-evaluation to the L&D function of the company. All participants agreed that meta-evaluation, in principle, was an important area of inquiry for the company as it works to develop a more comprehensive metrics and evaluation strategy. One respondent commented that it “serves as a catalyst to think through our evaluation strategy.” Another said that it “forces us to question what we really care about and then ask whether what we do in the program is helping us to arrive at those outcomes.” However, one of the respondents, while agreeing that meta-evaluation could be valuable, commented more specifically that this investigation was “not useful, because changes needed to have been made (to the program) three years ago - this is three years too late.”
The point around the timing of meta-evaluation and the actions that it might catalyze is an important one. In many ways, the value of this investigation (the meta-evaluation) will remain incomplete until this document or summary report is disseminated. It was clear to the investigator that the process of examining an already-completed evaluation through the lens of the four meta-evaluation standards was a useful one. The primary benefit was that it created the conditions for a conversation regarding the purposes and outcomes that a “successful” evaluation ought to pursue. The possibility of building an emerging consensus around evaluation priorities is an important foundation for building a comprehensive metrics and evaluation strategy. The meta-evaluation also underscored for the investigator the importance of creating alignment around expectations with key stakeholders before an evaluation and then having a clear communication plan in place to disseminate the findings of a given evaluation in a timely way to a comprehensive set of stakeholders. Ultimately, it is hoped that this meta-evaluation will serve as an important and appropriate first step of many that would be required to build a robust, comprehensive, and defensible metrics and evaluation strategy for the organization.

In terms of the positive outcomes of conducting the meta-evaluation, the process re-engaged a key set of stakeholders to reflect on the Success Case that had been conducted four years earlier. The value of this reflection was that it surfaced the “working theories” of these senior practitioners regarding what they viewed as important in evaluation. The passion and conviction from stakeholders that came out during the interviews were somewhat surprising to the researcher, and served to reinforce the importance of creating a space for these conversations. It underscored the reality that individuals appreciate being invited to share their views. To wit, each expressed views related to the propriety, accuracy, feasibility and utility of evaluation. More importantly, their views provided an important cultural “heat map” of what would
constitute a valuable evaluation. As a result, there were many suggestions in the interviews that
would serve to refine future inquiries relative to each of the meta-evaluation standards of the
evaluation approach in the organization.

As mentioned earlier, some participants expressed that the meta-evaluation was less
valuable given the time lapse between the original Success Case evaluation and this study. From
the perspective of the interviewer however, the time-lapse served to elicit more candid comments
relative to evaluation than may have been received immediately following the Success Case
evaluation. The perspective of time, and inevitable memory decay, seemed to serve as a helpful
filter ensuring that only the most important elements remained salient, and less consequential
elements were forgotten.

**Changes in Participants’ Views Over Time**

While none of the above findings is surprising, they are interesting. In the subjective
recollection of the investigator, the views expressed by participants in the study toward the
Success Case evaluation were less positive than when the study’s results were first shared. In
other words, the researcher remembered the participants in the study as having been less critical
and more positive about the Success Case evaluation than they appeared to be in the interviews.
One explanation for this is that perhaps there were high expectations of changes to the design of
the program as a result of the evaluation. While there were a number of concrete actions that
were taken as a result of the Success case evaluation, none could be considered either radical or
transformational.

A second explanation is that the more explicit invitation to review the evaluation report
more critically as part of this investigation served to overcome any organizational politeness
(social desirability) that might have prevailed four years earlier. Given that the investigator had
been responsible for the leadership program and sponsor of the Success Case evaluation, many of the participants may have felt inhibited in being more candid had the meta-evaluation happened earlier. Add to this that nearly two-thirds of the participants served as either faculty or coach for the program, they may have been more invested in a positive narrative.

A third explanation for the increased candor is that the continued evolution of the L&D function over the four years led participants in the investigation to raise expectations around what would constitute an effective evaluation, representing a form of “response-shift bias.” At the time of the Success Case evaluation, there was little formal work that had taken place around evaluation. In the following four years, however, a number of robust evaluations were sponsored, and so it is possible that the Success Case evaluation had less luster when compared to other work that had been done.

In addition, in the four years between the Success Case evaluation and the meta-evaluation, the organization continued to build out its leadership development curriculum. For example, the global leadership development program, which served as the evaluand for the Success Case, was the first leadership program in the nascent leadership core curriculum to be deployed globally, and represented the first formal application of the company’s Leadership Pipeline model. Given the importance of the program to the emerging Leadership Development strategy, the seniority of the participants, and the vision that this be a core global program, the stage was set to be ready for a more formal and comprehensive evaluation of the program (i.e., the Success Case evaluation).

In the four years following, the organization not only increased the annual delivery of the leadership development program by number of programs and participants by to roughly 600 participants per year, it also built and globally launched three additional programs as part of its
DISSERTATION: APPLIED META-EVALUATION

core curriculum. It also set out a five-year strategic plan to ensure that all forty thousand people managers in the company would participate in at least one of the core programs. One way to think about this change in organizational context and in the evolution of the Learning function is through the lens of a maturity model.

Changes in Participants’ Views and Organizational Maturity

Bersin & Associates have developed a four level taxonomy that they have called a Leadership Development Maturity Model (Mallon, Clarey, and Vickers, 2012). This model highlights a step-wise progression of an organization toward greater levels of maturity:

- Level 1 – Inconsistent Management Training – there is little or no management support for leadership development. Course offerings are not built around a strategic plan and are not progressive by level.
- Level 2 – Structured Leadership Training – the organization begins to focus on leadership skills and has defined a core set of competencies. Notable is the beginning of senior leaders to view leadership development as a priority and strategic imperative.
- Level 3 – Focused Leadership Development – the focus shifts to not only the individual leader but to the organization itself and its culture. There is more of a future orientation and also begins to incorporate a more blended approach.
- Level 4 – Strategic Leadership Development – leadership development is fully integrated with the overall talent management system and content is aligned with strategic priorities and is delivered through multiple channels.

Using this model as a guide, the launch of the Leadership Development program marks the beginning of the organization’s transition from Level 1 to Level 2. (This view was corroborated by the Learning function’s leadership team’s subsequent assessment of the
function, which was carried out at roughly the same time as the meta-evaluation). It follows that as organizations transition from one level to the next, the evaluation practices must also adapt and mature. For example, if a particular program is only offered a single time or is not part of a more strategic plan, the need for a thorough evaluation is likely to be far less than if a program is meant to serve as the anchor of an emerging global framework. Where there is a more stable and targeted investment in learning, it follows that the desire to understand impact and ROI would also grow. Thus, the Success Case evaluation from this vantage point was an appropriate and timely evolution in evaluation, and positive comments during the meta-evaluation seem to support this (e.g., “this was the most sophisticated evaluation I had experienced in the organization or elsewhere,” “this signals a more professional learning function,” etc.).

While it is impossible to know the degree to which of the above explanations, if any, influenced the views expressed relative to the utility of the Success Case evaluation, it is most likely that there was a combination of all three.

Role of the Investigator

There were a number of advantages and disadvantages to the role the investigator played as an insider in the organization. In terms of advantages, the most important of these was personal knowledge of the organization and key stakeholders. As an internal member of the company and part of the global leadership team of the function, I was keenly aware of organizational history, context, state of evaluation, the Leadership program, (which was the original evaluand), and the Success Case evaluation. It could be accurately said that there was no one in the organization with closer ties to this work, given the role I had played in the design and deployment of the program and subsequent sponsorship of the Success Case evaluation. This knowledge enabled me to identify key areas of focus and also to have a general sense of the
challenges facing the organization and those variables that would factor into a decision around the adoption of a given evaluation approach. The contacts and relationships that I had with the participants in the study were likely important factors in the high response rate for participation in the study. Personal recollections of actions taken and the overall process also served as additional inputs to go along with the data collected, including from the archival review.

There were also disadvantages to having been so embedded in the organizational system. Given the personal relationships that I had with the participants in the study, there may have been less candor in relation to their comments regarding the value of the Success Case evaluation or the meta-evaluation itself. On the other hand, having the trusting relationships may have led participants to be more candid, so this must remain an open question.

Another challenge to the research was the biases that I may have brought into the investigation. Given I was both personally and professionally invested in the perceptions of key stakeholders around the leadership program, the Success Case evaluation, and the meta-evaluation itself. It is probable that these biases influenced comments I made in the semi-structured interviews, where I sometimes wondered whether I was made clarifying summary statements, which may have inadvertently guided responses of participants. For example in the interview transcripts there are three occasions where I vocalize this concern and say “maybe I am leading the witness,” and in one case the interview responded with “yes, you are, but I agree with you.” A similar bias might have played itself out in the coding of the interviews. In this case, having another rater classify quotes was a means to counterbalance this bias. This was helpful, but it would not have been sufficient to fully overcome it, as I did not have the rater review all of the interview transcripts and classification of themes.
Limitations of the Study

There were a number of limitations to the study, which should be taken into account when reviewing the results, several of which have already been covered. First, the small sample size (n=21) was not enough to provide any testing of statistical significance for results from the quantitative survey. Given the exploratory nature of the study, the sample size was deemed adequate, but this limited the confidence in which any definitive conclusions might be stated. Similarly, to validate a meta-evaluation approach, applying it to a single evaluation provides no meaningful basis for comparison.

Second, by design, the investigator decided to include a less detailed evaluation report of the Success Case Evaluation to participants in the study versus a highly detailed technical report. This decision was based on the fact that the latter report had not been originally distributed to stakeholders at the time of the Success Case evaluations’ completion several years earlier. Similarly, the investigator felt that the evaluation report that was circulated was the better written of the two and was better suited for a broader audience and would require less time for review. In retrospect, however, it became clear in the interviews that the meta-evaluation participants would have preferred the more comprehensive and technical report given their sense of what they felt was “missing.” As noted earlier, this limited the extent to which the study was able to evaluate the SCM itself. In future studies with participants from within the HR function, it might be better to err on the side of providing as much technical information as is available. A shorter report, like the one circulated, could be used with stakeholders outside the function.

A third limitation to the study was that it was designed primarily with HR stakeholders in mind. As an initial exploration of meta-evaluation, this made sense given their awareness of the company’s leadership development programs and evaluation practices. Both the Success Case
evaluation and the meta-evaluation would have benefited from input from a broader set of stakeholders, most notably, from the business.

A final limitation was the fact that there was a single investigator who conducted the meta-evaluation inquiry. This made it impossible to replicate the speed with which a meta-evaluation would need to be completed to achieve maximum utility to the organization. Similarly, the fact that participants were aware that the investigation was being conducted by a colleague who was also serving as a student in a doctoral program may have diminished the sense of organizational importance of the investigation than if it had been, for example, mandated by a senior executive or Chief Learning Officer.

Contributions of the Study and Implications to the Field

Meta-evaluation is an area of increasing interest to the overall evaluation field, particularly for summative evaluations of programs where continued investment (e.g., in the form of government or endowment funded programs) is predicated on documenting their impact. As a result, more meta-evaluation activity has taken place in government and the public sector than in corporate settings. As mentioned earlier, there is a gap between the state of research and practice of evaluation corporate Learning functions. The gap is particularly pronounced in relation to leadership development programs, where the challenge of open transfer of learning is very high.

This investigation, while exploratory, represents an experiment in how meta-evaluation might be applied to evaluate the appropriateness of a method to evaluate a Leadership development program. The findings of the study indicated that there is a great appetite amongst learning practitioners in the organization to move beyond traditional Kirkpatrick Level 1 and 2 toward being able to credibly demonstrate the value that learning investments provide. The
strategic application of the meta-evaluation approach to existing evaluation processes would not only add value from the perspective of good hygiene, but create an opportunity to shape a thoughtful and more comprehensive metrics and evaluation strategy – one that would enable leaders of the learning function to be able to confidently report to sponsors and other key stakeholders on the value being produced for their investments.

In this investigation, the meta-evaluation had both summative and formative applications. As a summative evaluation, the Success Case Method applied to a leadership development program is promising as an approach to better apprehend the program’s impact. From the perspective of formative evaluation, the meta-evaluation provided insights that, if applied, would help to inform and refine the application of the Success Case approach within the company context. Not only did the meta-evaluation provide insight into how the Success Case approach might be more effectively applied, but it also generated information that could help the organization overall to strengthen its evaluation practices.

**Implications for Future Research**

At the time of this writing, the fields of corporate and higher education are undergoing a significant transformation with far-reaching implications. Technological advances (e.g., cloud computing, mobile devices, internet access, social networking sites, APIs, apps, wearable technology, etc.), have created new distribution channels for content that have opened possibilities for learning and development that were not imagined even ten years ago. While it is not the purpose of this research to explore these developments in any depth, it is worth noting a few of the more significant “disruptors” that are likely to continue to challenge assumptions around what constitutes training, who creates it, how it is accessed, and ultimately how it is evaluated. The advent of podcasts, TEDtalks, YouTube, MOOCs, and sites such as edX and
Kahn Academy are forcing L&D departments and formal educational institutions to fundamentally re-think their value propositions.

In the past, both corporate and formal education would take place in a classroom with an expert instructor. Today, these same instructors might deliver a lecture to a live classroom that is broadcast live or archived, which learners can access on any number of devices and locations. In this new world, corporate learning functions will be less focused on the design and delivery of content as they will be with curating the content and making it available to learners so that what they need can be accessed when they need it. As a result, learning is becoming less event-based and more continuous, less isolated, and more connected. In the case of Kahn Academy, learning has been “flipped” so that the lecture portions happen while children are at home, and the “homework” or application happens in the classroom where the role of the teacher is to provide feedback and individualized instruction. In this way, students are able to get just-in-time help as they need it, and the teachers are less focused on providing the lecture, but helping the students in the application. The opportunity to focus on immediate application because the learning meets a pressing need is particularly promising. Is there an opportunity to better track the impact of these actions as well?

How will measurement and evaluation of learning take place in the future? For one, it seems that organizations will need to determine how much to invest in tracking activities versus outcomes. How important will it be to stakeholders and sponsors of learning to know that an individual completed an online course, observed a specific TEDtalk or listened to a certain number of podcasts? With multiple, self-directed channels available for learning, how can impact be isolated? As learning becomes more continuous, is there an opportunity for evaluation to evolve to be more continuous? What are the implications of wearable technology as it relates
to reinforcing learning and tracking application? From an evaluation standpoint, how will what constitutes utility evolve in light of learning becoming more continuous? Given that technology enabled learning produces a lot of data in and of itself, will this make evaluation practices more feasible through access to data analytics? As noted before, evaluation of corporate learning has historically been inadequate, but will evaluation practice fall even further behind, or is there an opportunity to “leap-frog” over many of the current challenges into a more effective and robust set of practices?

In this time of rapid change and transformation it would seem that meta-evaluation can play a meaningful role as L&D departments experiment with different methods and measures to ascertain the value of training. Eric Reis (2011) has written about what he has termed Lean Start-up, which has as its aim to shorten product development cycles and minimize risk by adopting a combination of business-hypothesis-driven experimentation to quickly test ideas, learn from what works and what does not, and then “pivot” based on revealed insights. Meta-evaluation holds promise for this type of iterative experimentation. For example, L&D departments may want to compare the efficacy of different methods for the development of a similar competency area. The meta-evaluation framework would serve as a means to help determine which methods best serve the different modalities of the training, it may even provide a means for reaching more confident conclusions in terms of measuring training impact when comparing the modalities and associated trade-offs.

**Specific suggestions for future research.** To build on this investigation, a number of opportunities exist. The first would be to apply the meta-evaluation to evaluations of programs of the same genre. This would enable the Learning function to make a more informed decision around the appropriateness of an evaluation approach to that particular genre of training. Along
a similar vein, an evaluation approach could be applied to multiple genres of training. By conducting a meta-evaluation of the various evaluations, decisions could be made regarding which of the genres in question would be best suited to that particular evaluation approach. Given the results that emerged from this study and the time it has taken the researcher to complete it, there is a need to arrive at a more refined and targeted approach to meta-evaluation. For example, in the context of the company in question, the focus on accuracy and propriety could be minimized and this could be part of preliminary review in advance of conducting an evaluation. The question of feasibility also could be limited to those groups of individuals who participated in the evaluation itself. The use of focus groups could also serve as a more efficient means of obtaining and corroborating data versus the individual interview approach that was taken for this investigation.

In conclusion, as long as companies continue to invest in their employees’ development, there will be a need for evaluation in order to give a proper accounting for this investment. In the midst of tectonic shifts in education and what constitutes learning, meta-evaluation holds promise as a discipline that will enable corporate learning functions to be more strategic and credible in their efforts to determine and justify the value they deliver.
References


DISSERATION: APPLIED META-EVALUATION


DISSEETATION: APPLIED META-EVALUATION

AMACOM.


Retrieved from

http://www.cloacademy.com/articles/view/it_s_time_for_measurement_strategy/


*Advances in Developing Human Resources. 8*(3), 528 - 539.


*Advances in Developing Human Resources, 11*(8), 528-539.


APPENDIX A

Informed Consent Form

Study to Evaluate a Program Evaluation of a Leadership Program (Applied Meta-Evaluation)

You are invited to participate in a research study that I am conducting as part of my doctoral studies in Organizational Psychology at the Graduate School for Applied and Professional Psychology at Rutgers. This research study has the full authorization of our company and will be used to help us better understand how to evaluate our programs, and this study proposes a way to help us evaluate the evaluations that are used. In addition, it is hoped that the study will contribute to the overall training field as it evolves in its use of evaluation of corporate training programs.

Approximately 30 individuals (mainly employees of the company) will participate in the study, and each individual's participation will require a total of approximately two hours. The study procedures include completion of an online questionnaire which will be followed up with an interview (either over the phone or in person). I will also ask that you review the attached evaluation report in advance of completing the survey.

• In the online questionnaire you will be asked questions around your perceptions of the evaluation report. The questionnaire will take approximately 15 minutes to complete.

• A number of participants will be asked to participate in a second phase of the research which involves an interview where I will ask you to elaborate on some of the answers you provided in the online survey aimed at gaining insight and information regarding your views of the evaluation report and any relevant aspects of the study that may be relevant. This will take approximately 30 minutes.

If you agree to take part in the study, once the questionnaires and interviews have been completed, your name will be removed from any associated documentation and you will be assigned a code, so that all of your responses will remain anonymous. Your name will appear only on a list of participants.

There are no foreseeable risks to participation in this study.

Participation in this study is voluntary. You may choose not to participate, and you may withdraw at any time during the study procedures without any penalty to you. In addition, you may choose not to answer any questions with which you are not comfortable.

This research is confidential. Confidential means that the research records will include some information about you, such as your name and email address (linking to your code on the survey). Once the survey has been collected, the results will be kept in a password protected file.
DISSERTATION: APPLIED META-EVALUATION

and your email address will be deleted. The interview transcripts will also only have the coding on the final copy.

The Institutional Review Board at Rutgers University and I will be the only parties that will be allowed to see the data, except as may be required by law. If a report of this study is published, or the results are presented at a professional conference, only group results will be stated, unless you have agreed otherwise.

If you have any questions about the study procedures, you may contact me at xxx-xxx-xxxx. If you have any questions about your rights as a research subject, you may contact the Sponsored Programs Administrator at Rutgers University at:

Rutgers University Institutional Review Board for the Protection of Human Subjects
Office of Research and Sponsored Programs
3 Rutgers Plaza
New Brunswick, NJ 08901-8559
Tel: 848 932 4058
Email: humansubjects@orsp.rutgers.edu

You will be given a copy of this consent form for your records.

Sign below if you agree to participate in this research study:

Subject ___________________________ Date ___________________
Online Questionnaire

Please answer the following questions utilizing the following scale. If you feel you do not have enough information to provide an answer or if the question does not apply to you, please elect the option (6) – “do not know/not applicable,” After each item you will have the option of providing additional comments. Any additional information you provide will be useful to the study, but is not required in order to complete the survey.

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

Research Question #1: To what extent did the SCM meet the evaluation standards for utility in the context of the organization in which it was conducted?

*NOTE*: These bolded headings (above each set of questions) appear only for purposes of showing the organization of the questions under the research questions, but will not appear in the online questionnaire itself
1. To what extent did you feel the conclusions of the study were accurate?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

2. To what extent were the conclusions of the study clear?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

3. To what extent did you find the recommendations made for program improvement to be
relevant given the program and organizational context?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

4. To what extent was the study useful to you as it related to understanding the impact on the organization of the LC3 program?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:
5. To what extent did you find the study’s recommendations to improve the program to be actionable (e.g., possible to be implemented)?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

6. To what extent did you think that the recommendations suggested by the report, if implemented, would enhance the likelihood of participants applying the learning back on the job?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>
7. To what extent would you feel confident you could share the final evaluation report with different stakeholders (e.g., business partners, program sponsors, managers of participants) without a great deal of editing?

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

8. To what extent are you aware of any actions taken in response to the evaluation report (e.g., changes to program content, communications to participants, tools, etc.)?

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
</table>
Research Question #2: To what extent was the SCM a feasible approach for the organization in which it was conducted? (this serves as category, not survey item)

9. To what extent did you consider the evaluation to be cost effective?

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:
10. To what extent did the requirements for carrying out the SCM prove to be too time-consuming for participants in the study?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:   

11. To what extent did the requirements to carry out the SCM prove to be too time-consuming in relation to the value of the finding in the final report?

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>
12. To what extent did you find the delivery of the final report “timely” in the sense that the organization still had interest in the findings when the final report was distributed?

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

13. To what extent did you find the questions asked in the study to be free of anything ethically inappropriate?

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>To a little</td>
<td>To some</td>
<td>To a great</td>
<td>To a very</td>
<td>Do not</td>
</tr>
</tbody>
</table>

**Research Question #3: To what extent was the SCM a meet the evaluation standards related to propriety? (not an actual survey item, but category)**

13. To what extent did you find the questions asked in the study to be free of anything ethically inappropriate?
14. To what extent did you encounter any bias (e.g., cultural/racial/religious/gender) in the questions asked of participants?

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>To a little extent</td>
<td>To some extent</td>
<td>To a great extent</td>
<td>To a very great extent</td>
<td>Do not know/not applicable</td>
</tr>
</tbody>
</table>

Explain:

15. To what extent do you believe the researchers/research team maintained the confidentiality that had been promised? (Propriety – for personnel in the study)

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
</table>
1. To what extent were you aware of any potential conflicts of interest in the study that were not acknowledged or addressed? (personnel in the study)

<table>
<thead>
<tr>
<th>Not at all</th>
<th>To a little extent</th>
<th>To some extent</th>
<th>To a great extent</th>
<th>To a very great extent</th>
<th>Do not know/not applicable</th>
</tr>
</thead>
</table>

Explain:
2. To what extent were you aware of any potential conflicts of interest in the study that were not acknowledged or addressed? (personnel in the study)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain</td>
<td>Not at all</td>
<td>To a little</td>
<td>To some</td>
<td>To a great</td>
<td>To a very</td>
<td>Do not know/not</td>
</tr>
<tr>
<td></td>
<td>extent</td>
<td>extent</td>
<td>extent</td>
<td>extent</td>
<td>great extent</td>
<td>applicable</td>
</tr>
</tbody>
</table>

Explain:
Introduction (to be read): As you know, from my earlier email contact with you and the consent form which you kindly signed, you have been asked to participate in this study in order to provide information about your experience with the Success Case (Brinkerhoff) Evaluation of the Leading at ____ 3 program which was conducted in 2010. The purpose of this interview is to learn about your experience with and views of various aspects of the final evaluation report, and follow up on a number of item responses from the questionnaire you recently completed. In addition to this interview with you, I will also be interviewing a number of other key stakeholders who had varying degrees of involvement with the evaluation and the Leading at ____i 3 program. Our Learning function is interested utilizing this information to ensure that it continues to offer high quality leadership development programs and a more robust set of practices related to metrics and evaluation.

In addition to providing an evaluation of the Success Case Evaluation of the Leading at XYZ Bank 3 program, I am conducting this research study as part of my education in a doctoral program at the Graduate School of Applied and Professional Psychology at Rutgers University. The purpose of my dissertation is to gain an understanding of how Meta-Evaluation (evaluations of evaluations) might be used to better determine the merit and worth of a given evaluation method. As a result, we hope to more regularly and effectively understand the impact leadership development programs that are offered at the company. Do you have any questions about this?

Because this interview is part of the data collection process for my dissertation, I have asked you to sign a written consent form. As you know, I will also be recording our conversation in order to create a transcript for more effective review of the different interviews. That said, if at
any time, you would like any portion of our conversation to not be recorded, please let me know and I’ll stop the recording. If, afterwards, you are interested in reviewing the transcript, I will also make that available to you. Did you have any questions about the consent form, confidentiality, or how I’ll be using the data that is collected? [Review consent form, confidentiality, what will be done with data, etc.]

Now that we’ve covered the background of this evaluation, do you have any questions before we begin the interview?

Interview Protocol: The protocol will consist of asking questions in follow up to the online questionnaire responses to the questions below. Probing questions will be oriented toward gaining greater understanding of what was salient in the mind of the respondent when completing the questionnaire.

<table>
<thead>
<tr>
<th>Main Questions</th>
<th>Additional Probing Questions</th>
<th>Clarifying Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(areas of inquiry to be based on the questionnaire items)</td>
</tr>
<tr>
<td>• What relationship, if any, have you had to the Leadership 3 program?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Can you explain in what way you were connected to the original study of which you read the final report?</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>• Overall, can you say</td>
<td>• What made this so</td>
<td></td>
</tr>
</tbody>
</table>

<p>|</p>
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>what you found most valuable about the evaluation study?</td>
<td></td>
</tr>
<tr>
<td>• Overall, what did you find least valuable about the evaluation study?</td>
<td>• What made this not as valuable as it could have been?</td>
</tr>
<tr>
<td>• What would you say was the “utility” or level of usefulness of the study to our organization?</td>
<td>• Are you aware of any specific actions that were taken by the organization as a result of the evaluation? Which ones?</td>
</tr>
<tr>
<td>• What sense do you have about how feasible it was to conduct the study?</td>
<td>• Do you have any other evaluations to which you would compare it?</td>
</tr>
<tr>
<td>• As you think about how the evaluation study was conducted were there any areas that gave you discomfort?</td>
<td>• Ethical nature of the questions?</td>
</tr>
<tr>
<td>• Are you aware of any specific actions that were taken by the organization as a result of the evaluation? Which ones?</td>
<td>• Any biases to how this was approached?</td>
</tr>
</tbody>
</table>
Additional probing and clarifying questions were asked based on review of responses to the electronic survey items and related comments. Depending on the stakeholder group the participant represents, the probing questions will seek to elucidate more information. For example, in the case of a respondent being part of the study personnel, probing questions can seek more detail about “how” the study was carried out and also gain more information about any challenges that were encountered in the study. Also, where a response tends to differ from the general response pattern, clarifying questions will be asked to try and understand the thinking behind the response.

**Concluding Remarks:** Those are all the questions I have for you. Do you have anything else you would like to me know?

Thank you for participating in this interview. The information you’ve provided has been very helpful and provides me with a better understanding of the impact of the Leadership Development Program. I will be interviewing other key stakeholders and colleagues over the next few weeks. Once I’ve completed all the interviews, I will compile the information gathered into a comprehensive feedback report to the leadership team. If you are interested, I would be happy to provide you with a copy of this report.
If you think of anything else you would like to discuss with me or have any questions, please don’t hesitate to contact me at the phone number or email address listed on the consent form you have. Again, I appreciate you taking the time to meet with me.
Learning Impact Map for Leadership Program

<table>
<thead>
<tr>
<th>Knowledge and Skills</th>
<th>Critical On-The-Job Applications</th>
<th>Individual/Department Key Results</th>
<th>Organizational Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(The particular learning outcomes that are most important for the trainee to acquire)</td>
<td>(The few most important ways this participant can use the learning to produce key results)</td>
<td>(The few most important job results that can help produce business goals and can be accomplished by applying the learning)</td>
<td>(The business impact this trainee can best contribute to by applying the learning)</td>
</tr>
<tr>
<td>Understand the key challenges – think more broadly</td>
<td>Move from a day-to-day focus (managing) to a future focus (leading)</td>
<td>Leaders are able to retain top talent through a willingness to offer developmental experiences and engage in an ongoing, coaching dialogue</td>
<td>Focus on our distinctiveness</td>
</tr>
<tr>
<td>Understand how to manage and lead in a matrix</td>
<td>Develop self awareness as leader: Assess personal leadership capabilities, manage risk factors and potential derailers. Build and act upon personal development plan.</td>
<td>Innovative, client and customer focused solutions are developed through cross-business networks and discussions</td>
<td>- Global network</td>
</tr>
<tr>
<td>Understand the challenges leaders face in making leadership transitions</td>
<td>Develop a coaching strategy: Actively develop and hold your directs accountable for coaching, motivating, and managing their team members.</td>
<td>A pipeline of talent is developed and managed more strategically and purposefully</td>
<td>- Emerging Markets</td>
</tr>
<tr>
<td>Understand the framework of a high performing team—identify steps to improve performance of your team</td>
<td>Lead a high-performance team: Assess and manage your teams’ capabilities, business performance, and productivity.</td>
<td>The organization begins to measure team performance; Teams show an increase in performance; Team development becomes a “pull” request from line leaders</td>
<td>- Innovation</td>
</tr>
<tr>
<td>Build capability in coaching—developing a strategy</td>
<td>Foster and leverage relationships: Demonstrate the organizational savvy to navigate in the company’s highly complex, matrixed structure; utilize informal networks to</td>
<td>Check-up 360’s are used consistently to track performance; Leader development plans are complete and robust</td>
<td>- Scope</td>
</tr>
<tr>
<td>Understand how to leverage relationships and build connections (conduct stakeholder analysis)</td>
<td>Your business sees greater differentiation in</td>
<td>Our leaders are holding difficult, well delivered coaching conversations</td>
<td>Focus on our clients</td>
</tr>
<tr>
<td>Develop the ability to engage and energize others through communication</td>
<td></td>
<td></td>
<td>Deliver one business</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Simplify complexity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Our business is Powered by people</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Create a Culture of responsibility</td>
</tr>
</tbody>
</table>
work effectively across boundaries.  
- **Inspire others to action**: Articulate the vision and direction for your team to align with business strategy; build and refine your leadership style and presence to influence and motivate others.  
- Hold tough conversations  
- Identify ways to streamline processes  
- Identify ways to improve products/services

| performance and talent ratings: The highest performing employees are getting the biggest rewards and recognition from leaders  
Managers of Managers demonstrate readiness for promotion more quickly than planned
List of documents sourced as part of Archival Review

1. Email correspondence between investigator and Program Manager to define scope of Success Case evaluation
2. Original Success Case Evaluation proposal from external evaluators
3. Work Order for Success Case Evaluation
4. Various drafts of the Learning Impact Map (final included in Appendix D)
5. Email communication to solicit participation in Success Case Evaluation
6. Success Case Electronic Survey
7. Success Case Electronic Survey results (Excel file)
8. Leadership Program Facilitator Guide
   a. 2009 Version (pre-Success Case Evaluation)
   b. 2010 Version (post-Success Case Evaluation)
9. 2010 Leadership Program Business Plan
10. Pre-program Manager Call PowerPoint slides with speaker notes (post-Success Case Evaluation)
11. Success Case Evaluation (version which was circulated for Meta-evaluation)
13. Email correspondence from Program Manager to key HR leaders regarding Success Case Evaluation findings
14. Follow-up communication plan from Program Manager to key stakeholders around results of the Success Case Evaluation and proposed actions to be taken
APPENDIX F

Guiding Questions for Archival Review

1. To what extent was the sample of individuals selected to participate in the SCM representative of the overall population of participants? (Propriety)

2. What was the amount of time required for the entire study (contracting through the production of the final report)? (Feasibility)

3. To what extent were participants appropriately informed of the purposes of the study? (Propriety)

4. To what extent were participants appropriately informed of the confidentiality of the study? (Propriety)

5. To what extent were participants informed regarding how the results would be used? (Propriety)

6. To what extent were participants in the study briefed regarding the study’s outcomes after delivering the final report? (Propriety)

7. To what extent was the data obtained for the study representative of the whole population being studied? (Propriety)

8. To what extent were the survey questions asked clear and unambiguous for the participants in the study? (Propriety)

9. To what extent was the methodology of the study clearly documented? (Accuracy)

10. To what extent were the survey questions linked to the stated objectives of the program? (i.e., did the questions reflect reasonable expectations related to the intent of the program) (Accuracy)
Select the best answer for each question.

1. What were your expectations as you began leading? (Select one)
   - I really had no specific expectations other than to participate and somehow gain from it.
   - I had some idea of what I might learn from it, but not much beyond that.
   - I was very clear about what new skills and knowledge I could gain.
   - I had specific objectives not only for what I would learn but also how I would apply it in my work.

2. How were your expectations set? (Select one)
   - Any expectations I had, came only from my own thinking or from the e-mail communications.
   - Any expectations I had, came only from conversations with others (HR or previous attendees).
   - My manager and I talked generally about how I might benefit, but we did not jointly set expectations.
   - My manager and I jointly discussed and set expectations.

3. Using the scale provided below, rate the extent to which you might have applied any learning from this training experience

<table>
<thead>
<tr>
<th>Applications</th>
<th>Implemented, got positive results</th>
<th>Implemented, but not sure of results yet</th>
<th>Have not implemented it, but plan to</th>
<th>Was already doing this</th>
<th>Will not be doing this (not part of my role)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use the feedback from the assessments to improve my leadership effectiveness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hold my direct reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>accountable for coaching,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivating and managing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>their team members</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Took action to increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>team productivity based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on application of the Team</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use informal networks to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>work effectively across</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boundaries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coach my directs leveraging</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the GROW model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communicate in a way that</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inspires others to action.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build a personal,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comprehensive development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plan that reflects key</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>program learnings</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shared my experience and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>development areas with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>others, and asked for their</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>assistance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. To what extent is there accountability for using this training in your work? (Select one)
   - I feel fully accountable for applying this learning; my manager and I jointly set expectations
   - There are some mechanisms in place for follow up, but they are not used consistently
   - Any accountability I feel for applying this learning comes only from my own determination
   - No one other than I really knows or cares if I apply this training

5. To what extent have you received follow up support and coaching from your manager after attending the program? (Select one)
   - I have received extensive and helpful follow up support and coaching
   - I have received some follow up support and/or coaching
   - I have received very little follow up support and/or coaching—and would benefit if I had more
   - I have received virtually no follow up support and/or coaching

6. To what extent have you received follow up support and coaching from your Table Facilitator after the program? (Select one)
   - I have received extensive and helpful follow up support and coaching
   - I have received some follow up support and/or coaching
   - I have received very little follow up support and/or coaching—and would benefit if I had more
   - I have received virtually no follow up support and/or coaching

7. Which statement below best characterizes your experience regarding this training? (Select one)
DISSERTATION: APPLIED META-EVALUATION

- I have learned something, used it to produce business impact, & have solid evidence to describe it
- I learned something, used it, and expect worthwhile results though none have been achieved yet
- It was a valuable refresher & motivated me to use it in ways that led to worthwhile results
- This training was mostly a reminder of what I already knew and was already doing
- While I may have learned something new, I have not put it to use yet
- This training did not cover anything new or useful

8. In which region are you based? (Select one)
   - EMEA
   - Lat Am
   - Asia Pac
   - North America

9. In which business do you operate? (Select one)
   - xxxxxx
   - xxxxxx
   - xxxxxx
   - xxxxxx
   - xxxxxx
   - xxxxxx

Name _______________________________________________________________

Email Address________________________________________________________

(Your name is needed only because we may follow up with some respondents by telephone to better understand their experiences. Your responses will only be seen by the evaluation team and are fully confidential.)