

ABSTRACT

MURPHY, DEENA MICHELLE. *Beyond Accountability: An Empirical Study of the Factors Associated with the Use of Evaluation for Organizational Learning in North Carolina's Nonprofit Sector.* (Under the direction of Denis O. Gray.)

Nonprofits, as a unique set of organizations that are mission and not profit driven, account for approximately 13 percent of employment in the U.S., and collective assets represent approximately 2.5 trillion dollars. Economic uncertainty and intensified competition for funding have led to an increased demand for resource accountability throughout the nonprofit sector. The public, government representatives, foundations, and other nonprofit funders need to ensure that the programs they fund have value for the public while volunteers, who devote valuable time to programs, want to know whether they really are making a difference in the lives of their community. This focus on accountability means that despite an increased emphasis in the evaluation field on the importance of organizations using and not shelving evaluation results, little research exists on the factors that facilitate nonprofits using evaluation results to improve.

This cross-sectional quantitative survey research uses a sample of 283 diverse nonprofits in North Carolina to examine nonprofits' evaluation practices to gain insight into the organizational factors that facilitate the use of evaluation by the nonprofit. Building on previous research, this study offers a more robust set of measures to operationalize the constructs of leadership evaluation characteristics, stakeholder engagement, organizational learning capability, evaluation implementation, and three different types of use: use for internal learning, for accountability, and for image building. Using multivariate analysis, this study explores the interconnections between these organizational factors and the different types of use.

Findings from the path analysis models indicate that while all the organizational characteristics are significantly associated with use of evaluation, there are differences based on how these factors impact the different types of use of evaluation, providing a focus for future evaluation capacity building initiatives. These results emphasize that how use of evaluation is operationalized and measured is critical to understanding what factors may facilitate a nonprofit's own use of evaluation. This study further provides insights into the organizational characteristics, such as operational budget and experience with evaluation, that moderate the relationships between leadership evaluation characteristics, stakeholder engagement, organizational learning capability, evaluation implementation, and use of evaluation by the nonprofit.

Beyond Accountability:
An Empirical Study of the Factors Associated with the Use of Evaluation for Organizational
Learning in North Carolina's Nonprofit Sector

by
DEENA MICHELLE MURPHY

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Psychology

Raleigh, North Carolina

2007

APPROVED BY:

Joseph Herkert

Mary B. Wyer

Craig C. Brookins

Denis O. Gray
Chair of Advisory Committee

BIOGRAPHY

Deena M. Murphy received her undergraduate degree from the University of Hull in England, in 1993. After working with English programs in the U.K., Greece, and Saudi Arabia, Deena came to the U.S. in 2000 and completed her master's and doctoral degrees at North Carolina State University. Deena is currently a Principal Research Associate at The Institute for Community-Based Research, which is the North Carolina office of the National Development and Research Institutes, Inc. Deena is also a part-time instructor with the Women and Gender Studies Program at North Carolina State University and a Board member of The Women's Center of Wake County. Deena's current research interests surround evaluation capacity building in social service organizations; understanding the organizational factors associated with improving the quality of the data collection process and making evaluation a more useful tool that helps improve organizations. Deena's dissertation work involved empirically studying the factors that facilitate implementation and use of evaluation for organizational learning in North Carolina's nonprofit sector. Results from this research have been disseminated to nonprofit leaders, funders and support organizations to gain greater insights into how to develop systems that will build the evaluation capacity of nonprofit and other social service organizations, leading not only to increased accountability, but also to enhanced organizational learning.

ACKNOWLEDGEMENTS

I am grateful to a large number of individuals who provided a wealth of emotional, intellectual, and financial support that enabled me to complete this process.

Initially, this research would not have been possible without the support of the Institute for Nonprofits, College of Humanities and Social Sciences, North Carolina State University, with financial support from the A. J. Fletcher Foundation and the Office of Extension, Engagement, and Economic Development at North Carolina State University.

I would like to say thank you to the staff at The Community Foundation of Western North Carolina, the John Rex Endowment, and the Z. Smith Reynolds Foundation, who supported the general idea of this project and provided contact information about potential nonprofit participants. More thanks must go to Jennifer London and Nikki Harris, who worked with me to update this participant list.

I would also like to thank the nonprofit executive directors who gave up their valuable time to pilot the survey and provide useful feedback: Kevin Brice, Triangle Land Conservancy; Anne Burke, Urban Ministries; Hope Hancock, SPCA of Wake County; Michael Lowder, Artsposure; Barry Porter, American Red Cross; Jeanne Tedrow, Passage Home; Jean Williams, The Women's Center of Wake County; Jennifer Tolle Whiteside, North Carolina Community Foundation; and Alice Watkins, Alzheimer's Association of Eastern North Carolina.

I would like to thank Sal Alaimo, Joanne Carman, Barbara Metelsky, and Hallie Preskill for sharing their expertise regarding evaluation capacity building among nonprofits,

and for providing suggestions that improved the design of this work, in addition to encouraging me onward. I look forward to ongoing collaborations with you.

For consultation over statistical issues, I would like to thank Shevaun Neupert and Jason Allaire. Much appreciation also goes to Drew Rivers for helping define differences between the measures of use of evaluation.

For their ongoing support, encouragement, and humor, I would like to thank my work colleagues: Jill Anderson, Marge Cawley, Glenda Clare, Gail Craddock, Bob Hubbard, Doris Leaston, and Lillian Robinson. Special thanks must go to Gail for providing a very knowledgeable pair of SAS eyes and to Mindy McNeely, who—when I was at a loss why my path analysis model would not work—introduced me to Stephen Ward, who reminded me that sometimes “SAS is just stupid.”

To my wonderful committee members, Craig Brookins, Denis Gray, Joseph Herkert, and Mary Wyr; all of whom have uniquely contributed to my intellectual refinement, professional development, and personal growth, I say thank you. I would not have started nor reached the end of this journey without your knowledge, support, and guidance.

TABLE OF CONTENTS

LIST OF TABLES.....	x
LIST OF FIGURES	xii
I: INTRODUCTION AND OVERVIEW	1
Outline of paper.....	1
II: BACKGROUND AND CONTEXT	3
Introduction.....	3
Nonprofit Organizations.....	4
<i>Defining Characteristics of the Nonprofit Sector</i>	4
<i>Impact of the Nonprofit Sector</i>	5
<i>Taxonomy of Nonprofits</i>	6
<i>A Brief History of the Nonprofit Sector</i>	8
<i>Organizational and Environmental Characteristics Impacting Nonprofits</i>	11
<i>The Shift toward Accountability for Nonprofits</i>	12
<i>The Nonprofit Sector's Mixed Response toward the External Push for Accountability</i>	15
<i>Summary of the Nonprofit Sector</i>	19
Changing Evaluation Landscape.....	21
<i>The Transformation of Evaluation: from "Norm" to "Anomalous"</i>	21
<i>A Review of Influential Models and Approaches toward Evaluation</i>	25
<i>Summary of the Changing Evaluation Landscape</i>	31
The Link between Evaluation and Organizational Learning.....	32
<i>Theoretical Frameworks for Organizational Learning</i>	33
<i>Linking Organizational Learning and Evaluation</i>	35
<i>Summary of the Theory around Organizational Learning and the Linkages with Evaluation</i>	38
Using an Innovation Lens to Clarify how Evaluation can be Implemented and Used in the Nonprofit Sector.....	40
<i>Description of the Innovation Process</i>	41
<i>Social Innovation Model</i>	46
<i>Fidelity and Core Characteristics of Evaluation Implementation and Use</i>	47
<i>Factors that Impact the Innovation Process</i>	49
<i>Innovation Characteristics for Implementation</i>	50
<i>Summary of the Innovation Process as it Relates to Evaluation</i>	52
Summary: Toward a Conceptual Framework for Evaluation for Organizational Learning in the Nonprofit Sector.....	54

III: LITERATURE REVIEW.....	57
Introduction.....	57
Implementation and Use of Evaluation.....	58
<i>Definitions of Implementation.....</i>	59
<i>Definitions of Use and Theoretical Frameworks around Factors that Impact Evaluation Use.....</i>	60
<i>Summary of Definitions of Use and Theoretical Frameworks.....</i>	65
<i>Empirical work on Evaluation Implementation.....</i>	66
<i>Summary of the Theory and Research around Implementation of Evaluation...</i>	68
<i>Empirical Literature on Evaluation Use: 1971-1996.....</i>	70
<i>Summary of Empirical work on Evaluation Use Prior to 1996.....</i>	74
<i>Review of Empirical Work Since 1996.....</i>	75
<i>Research using Meta-evaluation.....</i>	82
<i>Case Study Approaches.....</i>	83
<i>Mixed Methods Research.....</i>	86
<i>Related Research: Knowledge utilization.....</i>	87
<i>Summary of Recent Research around Utilization of Evaluation and/or Knowledge.....</i>	88
<i>Use of Evaluation: A Critique of the Research and Implications for Nonprofits.....</i>	90
Organizational Learning.....	92
<i>Empirical Research around Organizational Learning.....</i>	93
<i>The Dimensions of the Learning Organization Questionnaire.....</i>	94
<i>The Assessing Learning Culture Scale.....</i>	101
<i>Summary of the Research on Organizational Learning.....</i>	102
Stakeholder Engagement	104
<i>Stakeholder Theory</i>	104
<i>Empirical Research</i>	106
<i>Summary of the Theory and Research around Stakeholder Engagement.....</i>	115
Role of Supportive Leadership	116
Summary of Empirical Research surrounding the Factors Associated with Implementation and Use of Evaluation for Organizational Learning: Resulting Research Questions and a Proposed Model for the Nonprofit Sector.	121
<i>Organizational Factors: Correlations</i>	122
<i>Multivariate Relationships</i>	124
<i>Mediating Factors</i>	126
<i>Moderating Factors</i>	127
<i>Summary of Research Questions and Proposed Model</i>	128
<i>Contributions to Field.....</i>	130

IV: METHODOLOGY	131
Research Design	131
Process	131
Survey Development	132
Participants	133
Measures	135
<i>Organizational Characteristics: Covariates</i>	136
<i>Dependent Variable: Use of Evaluation</i>	137
<i>Independent Variables</i>	142
<i>IV1: Leadership Evaluation Characteristics</i>	143
<i>IV2: Stakeholder Engagement</i>	146
<i>IV3: Organizational Learning Capability</i>	151
<i>The Organizational Learning Capability Scale</i>	151
<i>Mediating factor: Implementation of Evaluation</i>	157
Analyses	164
<i>Correlations</i>	164
<i>Multivariate Relationships</i>	165
<i>Moderators</i>	165
<i>Overall Model</i>	166
Dealing with Missing Data	167
V: RESULTS	169
.	
Research Question 1: Is there a Correlation between any of the Organizational Characteristics (Size, Age, Time doing Evaluation and Type of Nonprofit) and Use of Evaluation by the Nonprofit?.....	169
Research Question 2: Do the Dimensions Underlying Organizational Learning Influence Use of Evaluation by the Nonprofit?.....	170
Research Question 3: Does Increased Stakeholder Engagement Result in Increased Use of Evaluation by the Nonprofit?.....	171
Research Question 4: Do Leadership Evaluation Characteristics Predict the Use of Evaluation by the Nonprofit?	173
Research Question 5: What is the Combined Effect of Stakeholder Engagement, Leadership Evaluation Characteristics and the Dimensions Underlying Organizational Learning on the Dependent Variable of Use of Evaluation by the Nonprofit?.....	173
Research Question 6: Is there a Correlation between Implementation of Evaluation and Evaluation Use by the Nonprofit?	178
Research Question 7: Is the Combined Effect of Stakeholder Engagement, Leadership Evaluation Characteristics and the Dimensions Underlying Organizational Learning on the Dependent Variable of Use of Evaluation by the Nonprofit Mediated by Implementation?.....	180

Research Question 8: Are any of the Correlations (between Stakeholder Engagement, Leadership Evaluation Characteristics, the Dimensions Underlying Organizational Learning and Implementation) with Use of Evaluation by the Nonprofit Moderated by Organizational Characteristics, such as Size, Budget, Age, Funder, and/or type of Organization?	184
Overall Model.....	188
<i>The Measurement Model</i>	190
<i>The Initial Measurement Model</i>	190
<i>The Theoretical Model</i>	192
<i>Revised Model</i>	193
<i>Comparisons between the Three Measures of Evaluation Use</i>	198
<i>Internal Learning</i>	198
<i>Accountability</i>	200
<i>Image Building</i>	202
<i>Differences between Evaluation Use Measures</i>	204
<i>Moderating Impact of Budget</i>	205
<i>Small Budget</i>	206
<i>Medium Budget</i>	208
<i>Large Budget</i>	210
<i>Differences between Budget Groups</i>	212
<i>Moderating Impact of Time Doing Evaluation</i>	212
<i>Less Time Doing Evaluation</i>	212
<i>More Time Doing Evaluation</i>	215
<i>Differences between Time Doing Evaluation Groups</i>	217
VI: DISCUSSION	218
Review of Previous Findings.....	218
Review of Main Findings.....	220
<i>Evaluation Use</i>	221
<i>How Leadership Evaluation Characteristics, Stakeholder Engagement, and Organizational learning Capability Impact Different Types of Evaluation Use</i>	222
<i>Moderating Effects of Budget and Time Doing Evaluation</i>	225
Limitations.....	226
Future Research and Action.....	227
VI. REFERENCES.....	234
VIII. APPENDICES.....	254
Appendix A: Proposed Model of the Factors Associated with the Use of Evaluation for Organizational Learning in the Nonprofit Sector.....	255
Appendix B: List of Variables from Survey.....	256

Appendix C:	Cover E-mail Sent to Grantees by Executive Directors of Each Foundation.....	265 266
Appendix D:	On-line Survey Instrument.....	278
Appendix E:	Frequencies and Percentages for all Categorical Variables.....	279
Appendix F:	Descriptive Statistics for Numeric Variables and Scales.....	281
Appendix G:	Correlation Coefficients for Numeric Variables and Scales.....	282
Appendix H:	Graphs of Moderators.....	

LIST OF TABLES

Table 1.	Traditional Funding Approaches to Nonprofits	18
Table 2.	Evaluation Models and Approach	27
Table 3.	A Typology of Use of Evaluation Results.....	61
Table 4.	An Overview of Empirical Research on Use of Evaluation.....	79
Table 5.	Budget Comparison: Sample and North Carolina.....	134
Table 6.	NTEE Classification Comparison: Sample and North Carolina.....	135
Table 7.	Use of Evaluation Items, Factor Loadings on Each Scale, Variance Explained, and Coefficient Alphas.....	141
Table 8.	Leadership Evaluation Characteristics' items, Factor Loadings, Variance Explained, and Coefficient Alpha.....	145
Table 9.	Stakeholder Engagement Item, Groups, Factor Loadings, and Variance Explained.....	150
Table 10.	Organizational Learning Capability Items, Factor Loadings, and Unique Variance.....	156
Table 11.	Implementation of Evaluation Activities, Factor Loadings, Eigenvalues, and Variance Explained.....	161
Table 12.	Evaluation Activities Measure.....	163
Table 13.	Correlations between the Predictor Variables and Evaluation Use Measures.....	172
Table 14.	Factors Impacting Internal Learning.....	174
Table 15.	Factors Impacting Accountability.....	176
Table 16.	Factors Impacting Image Building.....	177
Table 17.	Correlations between the Implementation Variables and Evaluation Use Measures.....	179
Table 18.	Implementation Factors Impacting Evaluation Use.....	180

Table 19.	Factors Impacting Evaluation Activities and Evaluation Prevalence	182
Table 20.	Mediating Influence of Evaluation Activities (EA) and Evaluation Prevalence (EP).....	183
Table 21.	Moderators of Relationships between Predictor Variables and Evaluation Use Measures.....	185
Table 22.	Properties of the Measurement Model.....	192
Table 23.	Properties of the Theoretical and Revised Model.....	196
Table 24.	Properties of the Revised Model for Each Measure of Evaluation Use....	205
Table 25.	Properties of the Moderating Impact of Budget Models.....	210
Table 26.	Properties of the Moderating Impact of Time Doing Evaluation Models.....	215
Table 27.	Frequencies and Percentages for All Categorical Variables.....	278
Table 28.	Descriptive Statistics for Numeric Variables and Scales.....	279
Table 29.	Correlations Coefficients for Numeric Variables and Scales.....	281

LIST OF FIGURES

Figure 1.	The breakdown of the national nonprofit sector by operating budget.....	6
Figure 2.	The breakdown of the national nonprofit sector by NTEE designation...	8
Figure 3.	A model of the five stages in the innovation-decision process.....	42
Figure 4.	Interconnected multi-leveled processes of technological innovation.....	44
Figure 5.	Chain-linked model showing paths of information and cooperation.....	45
Figure 6.	A theoretical model of evaluation utilization.....	64
Figure 7.	Seven actions that promote organizational learning.....	95
Figure 8.	A hub and spoke model of the nonprofit stakeholders.....	105
Figure 9.	The dynamic links between stakeholder participation and utilization....	110
Figure 10.	The role of the leader in building capacity for program evaluation.....	119
Figure 11.	Survey questions related to use of evaluation.....	139
Figure 12.	Survey questions related to leadership evaluation characteristics	144
Figure 13.	Survey questions related to stakeholder participation.....	148
Figure 14.	Survey questions related to organizational learning capacity.....	154
Figure 15.	Survey questions related to evaluation activities.....	159
Figure 16.	Survey questions related to evaluation prevalence.....	160
Figure 17.	Overall path analysis model.....	189
Figure 18.	Revised path analysis model.....	197
Figure 19.	Revised path analysis model for Internal Learning.....	199
Figure 20.	Revised path analysis model for Accountability.....	201
Figure 21.	Revised path analysis model for Image Building.....	203

Figure 22.	Moderating Impact of Small Budget.....	207
Figure 23.	Moderating Impact of Medium Budget.....	209
Figure 24.	Moderating Impact of Large Budget.....	211
Figure 25.	Moderating Impact of Less Time Doing Evaluation.....	214
Figure 26.	Moderating Impact of More Time Doing Evaluation.....	216
Figure 27.	Effect of Evaluation Activities, conditional on Budget, on Internal Learning.....	282
Figure 28.	Effect of Openness and Experimentation, conditional on Budget, on Accountability.....	282
Figure 29.	Effect of Openness and Experimentation, conditional on Budget, on Internal Learning.....	283
Figure 30.	Effect of Systems Perspective, conditional on Budget, on Image Building	283
Figure 31.	Effect of Leadership Evaluation Characteristics, conditional on Budget, on Internal Learning.....	284
Figure 32.	Effect of Leadership Evaluation Characteristics, conditional on Time Doing Evaluation, on Image Building.....	284
Figure 33.	Effect of Openness and Experimentation, conditional on Time Doing Evaluation, on Image Building.....	285
Figure 34.	Effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation, on Internal Learning.....	285
Figure 35.	Effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation, on Accountability.....	286
Figure 36.	Effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation, on Image Building.....	286
Figure 37.	Effect of Evaluation Activities, conditional on Time Doing Evaluation, on Internal Learning.....	287

I: INTRODUCTION AND OVERVIEW

Outline of Paper

The purpose of this dissertation is to critique the research surrounding nonprofits, evaluation, and organizational learning and then contribute findings from an additional study, using a sample of North Carolina's nonprofits, in an attempt to identify the interconnections between organizational characteristics, including organizational climate and leadership, and implementation and use of evaluation for organizational learning rather than simply external accountability. While there is much discussion of the importance of nonprofits conducting and using evaluation for organizational learning, there are limited frameworks from which to operationalize and assess the factors that impact nonprofits using evaluation for organizational learning, rather than complying with external requirements.

In order to start to develop a framework, this dissertation will provide an overview of the background and context of the changing evaluation landscape in the nonprofit sector and then critique the multiple factors that impact nonprofits' implementation and use of evaluation in general. The concluding sections will make some observations on the existing empirical literature, offer the results from a new study and discuss the implications of this research with some suggestions for future investigations. While the results from this research will provide insights into current nonprofit evaluation practices and the function of evaluation, assessing how this differs by size, type and age of organization, the unique contribution will be related to the organizational factors that facilitate the internal use of evaluation by the nonprofit sector. The major research questions that will be asked include: How does the organizational context, characterized by factors such as leadership and

organizational climate, impact use of evaluation for organizational learning? How does level of stakeholder engagement in the evaluation process impact use of evaluation by programs and organizations? How do level of stakeholder participation and organizational climate interrelate and impact use of evaluation by programs and organizations? While these questions will not be definitively answered, the goal is to provide a useful foundation for future research into issues of evaluation and organizational learning in the nonprofit sector.

II: BACKGROUND AND CONTEXT

Introduction

To explore the context in which the nonprofit sector encountered the external demand for accountability and understand the impetus toward a changing evaluation landscape that promotes organizations using evaluation internally, it is necessary to provide an overview of literature from multiple fields.

As this research is focused on factors associated with the use of evaluation by the nonprofit sector, it is important to examine the role of the nonprofit sector in society. Highlighting the ecological context in which nonprofits have developed a greater level of responsibility for providing public services underlines how this is interlinked with a more general increased emphasis on resource accountability and performance.

To this end, it will be useful to ascertain the unique characteristics of the nonprofit sector, look at how taxonomies have been developed, describe what has led to the changing evaluation landscape across the sector and emphasize variables that should be considered in any nonprofit research. In addition, to better understand the evaluation process, it will be helpful to briefly review the historical context in which evaluation became “valued” and increasingly professionalized and look at essential aspects of different evaluation approaches. While this research is focused on the nonprofit sector, it will be useful to look at how for-profit organizations have approached evaluation within their organizations and whether this literature emphasizes any factors that may facilitate the move beyond accountability toward the use of evaluation by the nonprofit sector. These literatures will be compared to articulate any overlapping factors that should be looked at empirically.

Nonprofit Organizations

Strichman (2005) characterizes the nonprofit sector as a unique field where a spirit of altruism provides the motivation in the face of minimal resources. To explore this uniqueness, this section will define the characteristics of the nonprofit sector, provide a brief history of the context in which the nonprofit sector emerged, identify the current status of the nonprofit sector, highlight the organizational and stakeholder issues that are creating an impetus toward accountability and look at how the nonprofit sector is responding to this external push for evaluation. While an in-depth overview of the socio-cultural-historical context of the philanthropy sector is beyond the scope of this paper, insights from the literature will be drawn upon to offer some perspectives into the added complexity of the context of the nonprofit sector in regards to using evaluation.

Defining Characteristics of the Nonprofit Sector

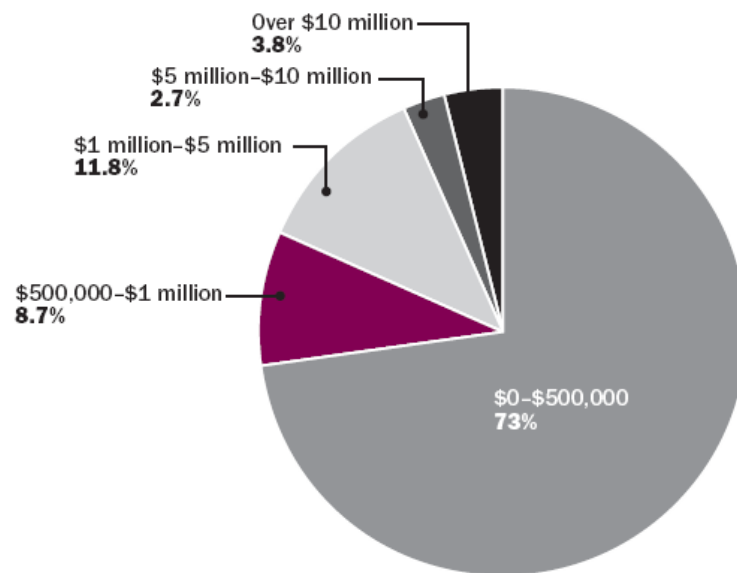
In his overview of America's nonprofit sector, Salamon (1999) attributes six distinct characteristics to nonprofits: they have a well defined public contribution; they are formal organizations and have been institutionalized to some degree; they are legally recognized and incorporated;¹ private non-governmental organizations; "nonprofit distributing," meaning they reinvest all profits into the organizational mission; self-governing, usually by an appointed board; and voluntary in nature, as volunteers play a key role in staffing and governing nonprofits. While many nonprofits distribute public benefits outside of their state, they are generally governed by state laws. This can result in discrepancies in relation to legal

¹ If non-religious nonprofits' revenues are above \$5,000, they must register with the Internal Revenue Service (IRS); if their revenues exceed \$25,000, they must file a version of an IRS Form 990.

privileges and liabilities. On a federal level, all nonprofits are exempt from federal income tax and a specific set of nonprofits can receive tax-deductible contributions according to Section 501 (c) (3) and (4) of Title 26 of the U.S. Tax Code.

Impact of the Nonprofit Sector

The current nonprofit sector is immense and accounts for approximately 13 percent of total employment in the U.S. The collective assets of these public charities represent approximately 2.5 trillion dollars (The Urban Institute, 2006). Using the 2004 IRS Business Master File, the National Center for Charitable Statistics (NCCS) at the Urban Institute identifies 1,397,263 distinct nonprofits in the U.S. (Urban Institute, 2006). This is a 29 percent increase from the 1,084,897 organizations that were in existence in 1994 (Urban Institute, 2004). Public charity 501(c) (3) organizations represent roughly half of this total population and rose 54% from 535,888 in 1996 to 822,817 in 2004. While densely populated large states such as California and New York have collectively more than 80,000 nonprofits with assets of more than one trillion dollars, smaller states such as Wyoming and North Dakota still collectively have around 3,500 nonprofits with assets of more than 9 billion. While nonprofits vary in size, as Figure 1 shows, the majority of nonprofits (73%) have budgets of under \$500,000.



Source: National Center for Charitable Statistics, Urban Institute (2003).

Figure 1: The breakdown of the national nonprofit sector by operating budget.

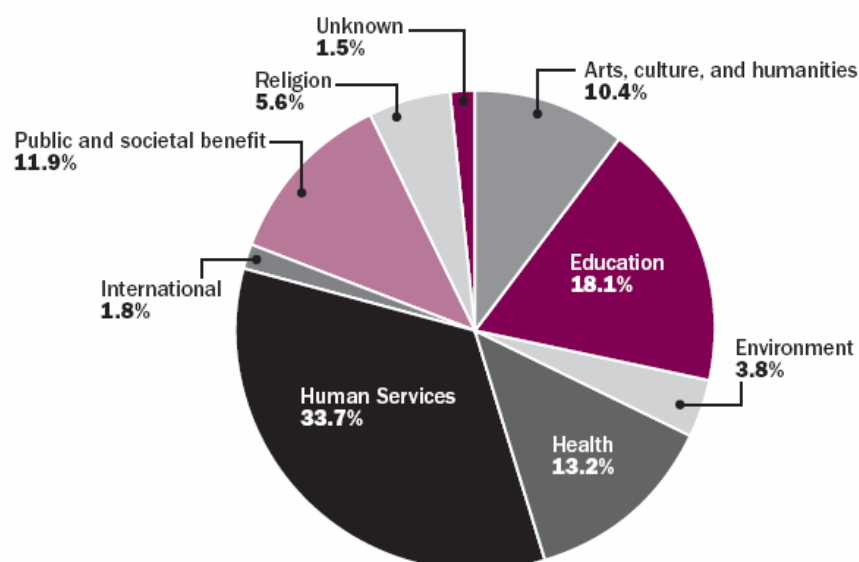
An Independent Sector report notes that the value of volunteer time within nonprofit organizations increased from \$169.6 billion in 1989 to \$225.9 billion in 1998. In 1997, more than 20,000,000 hours of volunteer time was donated by 55.5% of Americans (INDEPENDENT SECTOR, 2005). As an Urban Institute report notes in 2006, while the nonprofit sector currently represents approximately 13 percent of total employment in the U.S., there is no evidence that this trend will slow down and the sector continues to grow.

Taxonomy of Nonprofits

The most prevalent nonprofit typology, developed by a team of nonprofit scholars and practitioners in the 1980s and used by the Internal Revenue Service (IRS), the National Center for Charitable Statistics (NCCS), and many other research data organizations, is the National Taxonomy of Exempt Entities (NTEE) (Sumariwalla, 1986; Lampkin, Romeo, & Finnin, 2001). Based on the organizational purpose, the NTEE identifies ten main categories

of nonprofits that can be disaggregated into twenty-six major groupings and ultimately into about four hundred and fifty categories: arts, culture, and humanities; education; environment and animals; health and hospitals; human services; international and foreign affairs; public and societal benefit; religion-related; mutual membership/benefit organizations; and unknown.

As Figure 2 shows, in 2003, the Urban Institute found that the most common nonprofit groupings were: human service (34%), education (18%), health (13%), public and societal benefit (12%), and arts and culture (10%). One weakness of this coding system, highlighted by Lampkin et al., (2001) is that it does not always fully reflect the organization's programs, services, and activities. The example offered by Lampkin et al. (2001) is that of the 100 Black Men of America, Inc., which was founded in 1963 to "improve the quality of life of our communities, and enhance educational and economic opportunities for all African Americans" (<http://www.100blackmen.org>). According to the NTEE, this organization would be classified as a men's service organization (S82), yet the organization also provides programs for youth and the economic empowerment of minorities; this is not reflected in the NTEE code. Despite this weakness, NTEE remains the most prevalent coding system used by nonprofit researchers and is also used across all major nonprofit databases, including the NCCS.



Source: National Center for Charitable Statistics, Urban Institute (2003).

Figure 2: The breakdown of the national nonprofit sector by NTEE designation.

A Brief History of the Nonprofit Sector

In the United States, the voluntary sector can be traced back to the American Revolution and a constitutional settlement completed by the U.S. Supreme Court under the direction of Chief Justice Marshall (Hall, 2003; Hammack, 2001²). The nonprofit sector, classified by its organizational tax, policy, and regulatory purpose, came into being through Section 501 (c) (3) and (4) of the 1954 Internal Revenue Code (Hall, 2003). This allowed private organizations and trusts formed for charitable, educational, religious, and civic purposes to be exempt from taxation and allowed donors to make tax-deductible contributions. More broadly, Section 501 (C) (3) afforded organizations such as political parties, trade associations, and mutual benefit organizations various degrees of exemption and donor tax-relief, and constrained their surplus distribution (Hall, 2003). Notably, prior to

² These two papers (Hammack, 2001; Hall, 2003) provide a detailed overview of the history and development of the U.S. nonprofit sector.

1960, these nonprofit organizations predominantly served wealthier, more organized Protestants, Catholics, Jews and men and did little to empower those who were poor, people of color, or female (Hammack, 2001).

From 1956 to 2000, the nonprofit sector expanded from representing 2.8% of the domestic economy to 9% (Hammack, 2001). Hammack attributes this increase to three early factors: increased affluence among the U.S. population, the Great Societies programs initiated by President Lyndon B. Johnson and the Civil Rights movement. Increased affluence moved Americans to a more service-oriented society, as the share of gross domestic product devoted to services increased from 8% in 1950 to nearly 20% in 1990 (Hammack, 2001). The Great Societies program resulted in increased and ongoing federal subsidies, which grew from 0.4% of gross domestic product in 1962 to 4.44% in 1997 (Hammack, 2001). The Civil Rights movement played a pivotal role in the increase in the nonprofit sector as it bridged racial barriers within (northern and southern) Congress enabling support for federal funding of health care, social services, and education. In addition, to overcome resistance to racial integration at the local or regional level, Congress enabled Great Societies' programs to pay for private as well as government service providers (Sundquist, 1968; Hammack, 2001). The Civil Rights movement also upheld First Amendment rights to organize and set up nonprofit organizations that offered services to diverse societal needs, ensuring greater representation for the poor, people of color and women (Hammack, 2001; Silber, 2001).

While, historically, the liberal interpretation of the nonprofit sector revolved around the belief that alleviating poverty and improving social welfare required changes in social

and economic conditions; more recently, there has been an important changing philosophy around the nonprofit sector that can be greatly attributed to politically conservative strategists (Hall, 2003). In contrast to the liberal emphasis on the structural changes needed to alleviate poverty, the conservative perspective focused on the value and behavior of individuals as the locus of change. Hall (2003) builds on Hammack's (2001) early influences and points to the defeat of Barry Goldwater in 1964 as a pivotal point where political conservatives, who had traditionally viewed the growing nonprofit sector with suspicion, embraced the nonprofit sector as a way to create a counter-establishment to dominant liberal policies, using sympathetic policy research institutes, foundations and advocacy groups. Simultaneously, the recent growth of the nonprofit sector can be attributed to the conservative ideology of the 1980s, which provoked debate around federal and state responsibilities and devolution (Carman, 2005).

Reagan framed the significant social welfare budget cuts in the 1980s as a way to empower communities, private initiatives and states to have more control over pertinent social issues. George Bush Senior's 1988 Republican nomination speech took this devolution process one step further when he denounced the role of government in social welfare and envisioned "a thousand points of light" where voluntary, community-based initiatives would provide much needed social services. According to Hall (2003), this aggressive privatization of human services and devolution of government responsibilities to states and localities was simply a continuation of the privatization and devolutionary dynamics that had been building since the 1940s. Regardless, it significantly increased the

role that nonprofits played in the provision of social services and ironically, made nonprofits increasingly dependent on the government as a funding source (Carman, 2005).

Organizational and Environmental Characteristics Impacting Nonprofits

Researchers have identified a few important organizational and environmental characteristics that theoretically impact the organizational functioning of nonprofits and should be considered in any studies related to the sector. As Carman (2005) notes, these include organizational size, organizational age, leadership, geographic service area, service field, and funding sources.

DiMaggio (1988) and Cook (1988) emphasize the differences between smaller and larger nonprofits. DiMaggio (1988) notes that larger organizations are more complex, transparent, have greater accountability demands from multiple constituents, higher operating costs, and a higher probability of unionization. Cook (1988) perceives smaller nonprofits as facing more resource constraints, more instability, higher rates of failure, and less managerial professionalism. In regards to implementing and using evaluation, size will clearly impact an organization's ability or capacity to conduct evaluation activities.

Due to the pivotal role of leadership in managing and directing nonprofits, the executive director of nonprofits has an important impact on the functioning of nonprofits especially in regards to decision-making processes (Burke, 2001; Gronjberg, 2001; Salamon, 1999). In regards to age, Galaskiewicz & Bielefeld (1998) characterize newer nonprofits as having less intellectual, financial, or social capacity to capitalize on strategic opportunities.

Geographic service area impacts nonprofits through environmental attributes such as population density, socio-economic conditions and community capacity.

Service field and funding source are inextricably linked and can considerably impact how nonprofits perform and the level at which they conduct evaluation. Nonprofits often rely on multiple funders to sustain their practices, including: government contracts and grants; foundation grants; individual funders; investment income; fees for service; and product sales (Carman, 2005). This varies significantly by field however, as mental health and developmental disability organizations major funding may come from state and federal sources, while community development organizations rely on a variety of funds at the federal, state, local and individual level (Gronjberg, 1993; Carman, 2001).

While limited empirical research addresses the importance of these characteristics and will be discussed in the following chapter, sufficient nonprofit scholars have highlighted the potential impact of size, age, leadership, geographic service area, field and funding sources to warrant their inclusion in any nonprofit related research.

The Shift toward Accountability for Nonprofits

This increased dependency on government funding heralded an era of increased monitoring and accountability for nonprofits. Recent acts such as the Government Performance and Results Act (GPRA) of 1993, the Personal Responsibility and Work Opportunity Reconciliation Act of 1996, and the Balanced Budget Act of 1996 require organizations receiving government funds to conduct internal performance evaluations. Wolch (1990) suggests that this new dependency creates a limited nonprofit sector, which

favors larger nonprofits with the resources to generate proposals for government grants, while limiting the ability of nonprofits to criticize governmental policies. This has been echoed by other nonprofit researchers, who see this “contracting” structure as having potentially negative consequences for service delivery, democracy, and civic participation, in addition to changing the character of nonprofits by “professionalizing” the field (Carman, 2005; Smith & Lipsky, 1993). Other complaints related to this contracting relationship emphasize the financial uncertainty of social programs, the differing and often conflicting rules across states and localities, the extensive evaluation requirements, coupled with significant imbalances of power (Bernstein, 1991; Carman, 2005; Gronbjerg, 1993; Kelman, 2002; Smith, 1993; Stone & Cutcher-Gershenfeld, 2001).

Other prominent nonprofit researchers, such as Salamon (1995) and Saidel (1989) take a different perspective and perceive the government/nonprofit relationship as more of an interdependent partnership that will benefit both parties. Salamon (1995) perceives this “third party government” system as offering nonprofits the autonomy of adapting government programs to fit the community context, allowing nonprofits to play a larger role in coordinating decentralized government policies and reducing the administrative issues emanating from large government bureaucracy.

In the move toward increased requirements on accountability for nonprofits, other events played a key role. Issues surrounding expenditures by The United Way of Greater Washington in 2002 and The Nature Conservancy’s suspicious land deals and loans to the board led to the media and some nonprofits scholars (Fleishman, 1999; Jeavons, 1994) questioning the integrity of nonprofits and emphasizing the need for increased accountability

and transparency to garner the public trust. This was compounded by fiscal pressures at the city and state level which raised questions as to the appropriateness of nonprofits' tax-exempt status and societal contribution (Brody & Cordes, 1999). Simultaneously, an increased focus on social entrepreneurship within the nonprofit sector increased the competition between for-profit and nonprofit organizations and, as nonprofits drew funding from multiple sources, both private and public, caused renewed focus on how "independent" the sector really was (Stoecker, 1997). These events led to the media emphasizing the need for increased transparency and accountability and an Independent Sector Panel on the Nonprofit sector in 2002 (Ebrahim, 2005a).

The for-profit sector further facilitated the nonprofit sector's move toward accountability through the widely publicized accounting scandals of corporations such as Enron and WorldCom. In 2002, Congress responded to these scandals with the passing of the American Competitiveness and Corporate Accountability Act (the Sarbanes-Oxley Act) that provided publicly traded companies with a set of standardized governance, financial, and auditing practices (Carman, 2005). While this legislation strictly governs the for-profit sector, it has caused other state attorneys generals and nonprofit leaders to advocate for similar adherence within the nonprofit sector (BoardSource & INDEPENDENT SECTOR, 2003; Carman, 2005; Jones, 2003).

Given the enlarged role of the government, the ongoing professionalization of the nonprofit sector, the improvements in data monitoring systems, and the increased emphasis on accountability and transparency within the for-profit sector, it seems somewhat inevitable that the nonprofit sector would encounter the *zeitgeist* for evaluation. In a Brookings report

on nonprofit management reform in 2000, scholar Paul Light writes that “a nearly unanimous consensus has emerged that non-profit organizations have to improve their performance” (p.13).

The Nonprofit Sector’s Mixed Response toward the External Push for Accountability

The emergence of the nonprofit sector as a key player in providing social services combined with accountancy scandals, changing legislation in the for-profit sector, economic uncertainty and intensified competition for funding have led to an increased demand for resource accountability throughout the nonprofit sector. The public, government representatives, foundations and other nonprofit funders need to ensure that the programs they fund have value for the public while volunteers, who devote valuable time to programs, want to know whether they really are making a difference in the lives of their community (Murphy-Medley, 2005).

This increased demand for accountability, both internal and external to the philanthropic sector, has put a great deal of pressure on nonprofits to evaluate their work, often at the expense of improving the effectiveness of their programs and building the capacity of their organizations (Plantz, Greenway, & Hendricks, 1997; Salamon, 1999). This emphasis on external accountability has been critiqued as creating an environment where oversight and measurement prevail over innovation and creativity, where failure is not perceived as an opportunity to learn, and where short-term measurable results are focused on to the detriment of mission improvement and longer-term social change (Ebrahim, 2005a; Christensen & Ebrahim, 2006). Nonprofits emphasize that they have multiple and competing

accountability demands, such as to funders, to their clients, and to their organizational mission, and too great an emphasis on oversight and measurement can result in a hesitation to examine failures even when they could lead to better outcomes and improve the program and/or organization's effectiveness. Consequently, it becomes important to find a balance between external transparency, results, and reporting and other processes such as innovation, critical reflection and internal (or organizational) learning (Christensen & Ebrahim, 2006)

Adding to this complexity is the ongoing emphasis on evidence-based practice, which promotes the use of programs and practices that have data behind them that can inform an organization's decision making (Institute of Medicine, 2001). This creates a tension within nonprofits that may have adopted programs and practices that have little or no data to support their effectiveness and few resources to develop a system that measures outcomes and monitors performance. Consequently, many nonprofits contend that they lack the capacity in terms of education, skills, training and resources to respond adequately to these accountability expectations. These nonprofits claim they lack the capacity to generate valid, reliable and meaningful data on their philanthropic work that may improve their program and organizational practices (Bozzo, 2000; Hatry, 1999; Newcomer, 1997; Panel on Accountability and Governance in the Voluntary Sector, 1999).

In an effort to respond to this era of accountability and transparency, the nonprofit sector is looking to emerging trends in the evaluation field, which is increasingly highlighting the importance of evaluation as a tool that can play a central role in mediating the relationship between accountability and internal (or organizational) learning (Plantz, Greenway, & Hendricks, 1997; Poole, Davis, Reisman, & Nelson, 2001; Porter & Kramer,

1999). Since the mid-1990s, major foundations and other intermediary organizations, such as The United Way, have been investing in evaluation programs and units to enable nonprofits to conduct evaluation. These funders have dedicated resources to improving grantmaking and measuring outcomes and funders such as Kellogg and The United Way have focused on providing evaluation guides, tools and resources that will make evaluation and performance measurement more accessible to nonprofit managers (Bozzo, 2000; Hatry, van Houten, Plantz, & Greenway, 1996). While these tools are useful guides to the technical details of the evaluation process, little evidence exists to document how nonprofits are conducting evaluation and how useful this information is to their organization. In addition, these efforts often overlook the organizational capacity or resources necessary to conduct an evaluation. Consequently, despite these efforts, little progress has been made to develop evaluation capacity in the nonprofit sector through education, training, and skill building (Bozzo, 2000).

Strichman (2005) further emphasizes the critical role that funders have played in creating a focus on programs to the detriment of developing sustainable organizations. As Table 1 shows, traditional funding approaches have permeated the nonprofit sector with a way of thinking that often does not recognize the organization behind the service. This is reinforced by Letts, Ryan, & Grossman (1999), who note that the nonprofit sector has a “culture of service” that creates mission driven organizations with an underinvestment in organizational capacity.

Table 1

Traditional Funding Approaches to Nonprofits

Traditional Funding Approach	Rationale of Approach	Unintended Negative Consequences
Funding programs and under-investing in organizational expenses or administrative costs.	Service-oriented: desire to ensure that the bulk of resources go directly to beneficiaries.	Organizations lack resources to invest in capacity building.
Emphasizing new, innovative programs or projects.	Fulfilling the unique ability of the funder to test programs and design social policy.	Organization can be pushed beyond their core mission as they continually “recreate” new programs to meet funder’s interests.
Short-term investments.	Ensures that grantees do not become dependent, while allowing funders the flexibility to fund a variety of projects.	Not conducive to long term planning of organizational capacity.
Hands off approach.	Reluctance to “interfere” in the work of the nonprofit.	Works against the creation of a partnership where nonprofits can expose weaknesses and work with funders to address organizational challenges.

Note. Adapted from Strichman, 2005, p.66

Historically, creating social impact through program success has taken precedence over long term investments in infrastructure to promote organizational development and create sustainable organizations. Limited resources create a tension between building organizational capacity and providing much needed services. While the for-profit sector emphasizes the importance of investing in organizational capacity, nonprofits are often mistrustful of adopting “business strategies” that could improve performance as they seek to differentiate themselves from the for-profit sector (Letts et al., 1999). The instability and uncertainty related to funding sources also complicate the institutionalization of management practices. As nonprofits constantly struggle to support their programs, the majority of time is spent providing services. As McHargue (2003) comments, overburdened, under-staffed nonprofits have little ability to invest in skill development, reward staff performance or provide financial incentives to engage in activities that may facilitate organizational development.

Summary of the Nonprofit Sector

The nonprofit sector offers a unique context in which to research the role of evaluation. Nonprofits are grounded in a sector of mission driven organizations with limited resources that historically undervalue the importance of investing in organizational capacity. As organizations, nonprofits have very distinct structural and legal characteristics and are generally governed by state laws. Theoretical research around organizational and environmental characteristics impacting the nonprofit sector emphasizes the importance of

including organizational size, age, leadership, geographic service area, field, and funding sources as variables in any research.

In 2004, the nonprofit sector accounted for 8.3 percent of the wages and salaries paid in the United States. From 1996 to 2004, the nonprofit sector grew by 29 percent from 1.1 million nonprofit organizations (nonprofits) to 1.4 million. Public charity 501(c) (3) organizations, which represent roughly half of this population, rose 54 percent from 536,000 in 1996 to 823,000 in 2004. Total assets of these public charities account for 2.5 trillion U.S. dollars. If we take into account volunteer time (109 million Americans, representing 9 million full-time employees), the nonprofit sector represents approximately 13 percent of total employment in the U.S. (Urban Institute, 2006). Reasons for this growth can be attributed to several historical factors, though nonprofit researchers largely point to the devolution of social programs from the federal government to the states and localities as having a major impact on program service delivery and the increase in the number of nonprofits (De Vita, 1999; Carman, 2005).

In addition to significantly expanding in size, the nonprofit sector has become increasingly dependent on government funding, which has promoted an increased emphasis on monitoring and accountability. This has caused major funders to invest in evaluation and provide much needed resources to make evaluation and performance measurement more accessible to grantee nonprofit managers. While this is a positive step forward, it again ignores the importance of investing in infrastructure to build the capacity of nonprofits to implement and use evaluation effectively. More importantly, this emphasis on accountability has been critiqued as creating a punitive environment that largely ignores the potential

benefits of an organization using evaluation internally and learning from successes and failures. To achieve this, the nonprofit sector is looking to emerging trends in the evaluation field, which is increasingly highlighting the importance of evaluation as a tool that can play a central role in mediating the relationship between accountability and internal (or organizational) learning.

Changing Evaluation Landscape

While there are a variety of definitions of evaluation, Russ-Eft and Preskill (2001) combined the common themes of these and defined evaluation as a “process of systematic inquiry to provide information for decision-making about some object—a program, project, process, organization, system, or product” (p.204). To understand the context in which the nonprofit sector encountered evaluation and how it can move beyond accountability, it is necessary to have some understanding of the changing evaluation landscape and the transformation of evaluation itself as a field. While an extended overview of the history of evaluation is not the purpose of this paper, a brief review of some of the context in which evaluation became “valued” as a means for providing important information that would assist in learning, decision-making, and action will be useful.

The Transformation of Evaluation: from “Norm” to “Anomalous”

A variety of evaluation researchers have suggested that evaluation dates back to Socrates, who introduced the idea of critical inquiry into the teaching and learning process, and/or the Chinese dynasties, with the use of testing procedures to elicit membership into specific areas of employment (Guba & Lincoln, 1981; Stufflebeam & Shinkfield, 1985;

Worthen, Saunders, & Fitzpatrick, 1997). Regardless, it is clear that the concepts and practice of evaluation have been used across various cultures for a long period of time.

Modern practices of evaluation are inextricably linked to the public education, mortality and health and military personnel testing (Russ-Eft & Preskill, 2001). Russ-Eft and Preskill (2001) attribute the public's dissatisfaction with educational programs as resulting in Horace Mann and the Boston Board of Education using essay questions to replace oral examinations and recitation in the 1840s. Quantitative surveys used to measure mortality and health in the seventeenth and eighteenth centuries became linked to worker efficiency and were used increasingly to boost industrial effectiveness in the early twentieth centuries (Taylor, 1911). Guba and Lincoln (1989a, 1989b) note the use of mental tests to screen and classify military personnel in World War I as pivotal in promoting the norm-referenced testing that accompanied university and school accreditation. Largely due to the efforts of Ralph Tyler at the University of Chicago in the 1930s and 1940s, norm-referenced testing became replaced with criterion-referenced testing and focused more on reaching planned objectives than test norms (Guba & Lincoln, 1981; Russ-Eft & Preskill, 2001; Worthen, Saunders, & Fitzpatrick, 1997). However, most military and educational testing continued to focus on standardized testing and the role of the evaluator clearly remained one of a technician with a focus on the robustness of the research design and statistics (Guba & Lincoln, 1989a, 1989b).

Russ-Eft and Preskill (2001) attribute the onset of modern-day evaluation to the Kennedy administrations inclusion of an evaluation clause to the passing of the Elementary and Secondary Education Act in 1965. This essentially required all programs using federal

grants to include an evaluation component. While these early attempts to measure the impact of social programs have been described as analogous to the California gold camps (House, 1993), important work by Donald Campbell (1969) helped to professionalize the evaluation process by advocating for experimental designs, which could be continually tested to determine effectiveness and impact on social change. Joseph Wholey's (1975, 1976, 1979) work expanded on this and caused evaluation practitioners to encourage program staff to question the logic of their programs and practices, bringing the concept of evaluation into the program's design and assessing the feasibility of program objectives, critiquing the "and then a miracle happens" approach of many social programs. These early efforts and growth in the need for evaluation practitioners led to the establishment of several evaluation centers (such as the University of Illinois at Urbana-Champaign, Western Michigan University), promoting a growth in evaluation research and theoretical perspectives on how evaluation should be approached.

The merging of the Evaluation Network, predominantly university professors and educational evaluators, and the Evaluation Research Society, predominantly government-focused evaluators, resulted in the formation of the American Evaluation Association in 1985 (Russ-Eft & Preskill, 2001). While the 1980s was impeded by the Reagan administration's cuts in social programs and reduction to the point of removal in the requirement of evaluation in the federal grant process, the field of evaluation was promulgated by the growth in internal evaluation units in a variety of governmental and non-governmental organizations (House, 1993; Russ-Eft & Preskill, 2001). This pattern of mixing internal and external evaluation continued in the 1990s, with a growing understanding of how evaluation could move beyond

simple accountability toward improving program and practice effectiveness, enhancing organization and employee learning, and improving services through greater understanding of resource allocation and service needs (Patton, 1997; Russ-Eft & Preskill, 2001).

In summary, while the evaluation field is clearly still growing and professionalizing and offers a variety of models and approaches which will be reviewed next, it has shifted from a external top-down statistically driven approach toward a pluralistic approach which encompasses myriad methods, measures, criteria, perspectives, audiences, and interests (Russ-Eft & Preskill, 2001, p.46). The *value* latent in *evaluation* is acknowledged and there is a growing literature of the politics and ethics surrounding the knowledge generated by the field.

A Review of Influential Models and Approaches toward Evaluation

Russ-Eft & Preskill (2001) do a thorough job of assessing what they perceive to be the eleven most influential evaluation models and approaches developed from 1960 to 2000 (pp.47-61). What has been added to Russ-Eft & Preskill's (2001) overview is the important contribution of feminist evaluation techniques. While a discussion of the specifics of all these approaches is not relevant to the current paper, Table 2 outlines each approach and describes: those most associated with this work; intended users of evaluation findings; degree of client or stakeholder involvement; underlying assumptions; primary methodology; and the major focusing question. These characteristics outline much of the ongoing tensions among evaluation researchers around how evaluations should be defined, how and who should be involved in the design and the process of implementation.

Using these models and looking at how evaluation can benefit the nonprofit sector, what seems important to discern are the core components of what makes the evaluation process successful in terms of quality, implementation, and use of findings to inform decision-making and improve programs, services, and organizations. As one core component of a successful evaluation process includes the importance of stakeholder involvement in evaluation, I will review aspects of those models rated "high" in levels of stakeholder involvement; notably utilization-focused evaluation, participatory or collaborative evaluation, empowerment evaluation and evaluation for organizational learning.

Michael Patton (1978, 1986, 1997) defines *utilization-focused evaluation* (UFE) as a "process for making decisions about and focusing an evaluation on intended use by intended users (1997, p.317). The role of the evaluator is one of a facilitator rather than an external

independent judge and UFE assumes that stakeholders will play a major role in all aspects of the evaluation process from decisions surrounding design through to how the findings are used. As can be seen from Table 2, the major focusing questions for UFE are “What are the information needs of stakeholders and what plans are there for using the findings?” Patton (1997) positions UFE as a user-centered evaluation process, which allows the evaluator(s) and other stakeholders to choose an evaluation design and data collection methods that fit the specific information needs rather than advocating a specific approach. Patton emphasizes that this often involves the use of qualitative methodologies but can also include a mixed-method approach, utilizing both qualitative and quantitative methods.

Table 2

Evaluation Models and Approach

Model/ Approach/ Year	Authors	Intended Users	Level of Stakeholder Involvement	Assumptions	Primary Methodology	Major Focusing Questions
Behavioral objectives (1930s-1960s)	Tyler; Bloom; Mager; Popham	Managers	Limited	All objectives prespecified; quantifiable outcomes	Achievement tests; performance data	Is the program achieving its objectives? Is the program producing outcomes?
Responsive (1967)	Stake	Organization/community members	Moderate	Stakeholders know what they need to know	Mixed methods; primarily case study using qualitative methods	What does the program look like to different people?
Expertise (1970s) and accreditation (1800s)	Eisner	Government agencies; professional boards; organizations members	None	Experts know what is good	Document review; interviews; observation	How would professionals rate this program?
Goal free (1973)	Scriven	Consumers	Limited	Criteria can be specified; goals and objectives are unknown to evaluator	Observations; document review; interviews	What are the anticipated and unanticipated effects?
Adversary/judicial (1973)	Owens; Wolf & Rosenberg; Levine	Government agencies	Limited	Balanced presentations of facts can be provided	Public hearings; mock trial	What are the arguments for and against the program?
Consumer-oriented (1974)	Scriven	Consumers	None	Standards of quality performance can be prespecified	Checklists; usability labs	Would an educated consumer approve of this program or product?
Utilization-focused (1976)	Patton	Community/organization members	High	Stakeholders know what they need to know—emphasis is on use of findings	Mixed methods, though often qualitative	What are the information needs of stakeholders and what plans are there for using the findings?
Participatory and/or collaborative (1987)	Cousins & Earl; King; Greene	Community/organization members	High	Grounded in democratic decision-making processes with no political agenda	Mixed methods, though primarily qualitative	What are the information needs of those closest to the program?
Theory-driven (1987)	Bickman; Chen; Smith	Government agencies	Moderate	The evaluator can help identify the underlying theory of programs	Mixed methods, though primarily qualitative	How is the program supposed to work? What are the assumptions underlying the program's development and implementation?
Empowerment (1993)	Fetterman; Mertens	Community members	High	Goal to empower stakeholders	Mixed methods, though primarily qualitative	What are the information needs to foster improvement and self-determination?

Table 2 (continued)

Model/ Approach/ Year	Authors	Intended Users	Level of Stakeholder Involvement	Assumptions	Primary Methodology	Major Focusing Questions
Organizational learning (1990s)	Preskill & Torres; Cousins & Earl; Owens & Lambert	Organization members	High	Organization is interested in learning from and using the evaluation findings	Mixed methods	What are the information and learning needs of individuals, teams, and the organization in general?
Feminist evaluation (1990s)	Sielbeck- Bowen, Brisolara, Seigart, Tischler, Whitmore	Community/ organization members; focus on oppressed	High	Grounded in feminist theory; explicit goal of social justice	Mixed methods	How can we authentically communicate diverse perspectives throughout the evaluation process?

Note. Adapted from Russ-Eft & Preskill, 2001.

Worthen, Sanders, & Fitzpatrick (1997) highlight that *participatory* and *collaborative* evaluation approaches were developed in response to mechanized external evaluator approaches that focused on local, program-level impacts and often had little understanding of how to measure and articulate the social impacts of the program(s) being evaluated. Instead, evaluators such as Cousins and Earl (1992, 1995), Greene (1987, 1988a), and King (1995) advocated the inclusion of those stakeholders with sound knowledge of the program in the evaluation process. This emphasizes the importance of understanding and articulating the complexities associated with the program and enabling stakeholders to use the findings to inform decision-making practices. Table 2 shows that the major focusing question surrounding participatory and collaborative evaluation revolves around the information needs of those stakeholders closest to the program. The assumption underlying this is that stakeholders can use data to engage in a democratic decision-making process that will enable them to reach consensus and understanding about the impact and effectiveness of their programs and practices (Russ-Eft & Preskill, 2001). As with UFE, participatory and

collaborative evaluation approaches do not advocate for a specific ideological or methodological approach but instead use a mixed methods approach. In recent years, a wide range of participatory and collaborative evaluation approaches caused Cousins and Whitmore (1998) to develop a classification scheme that identifies participatory and collaborative evaluations as purposefully pragmatic or transformative (Russ-Eft & Preskill, 2001). Practical evaluations are intended to inform decision-making and facilitate problem solving while transformative evaluations seek to empower less powerful community members and groups. While much of the underlying principles are shared, Cousins and Whitmore (1998) differentiate these two purposes by assessing control of the process, identification of stakeholders, and level of participation through the evaluation process.

While *empowerment evaluation* is clearly rooted in the principles underlying community psychology, it is largely associated with David Fetterman (1994, 1996). Fetterman (1996) defines empowerment evaluation as utilizing the tools of evaluation to “foster improvement and self determination” (p.4). Fetterman (1996) believes that empowerment evaluation can be differentiated from other participatory approaches as it has an explicit political agenda of empowerment and a bias toward the “disenfranchised, including minorities, disabled individuals, and women” (p.27). As Russ-Eft & Preskill (2001) note, this type of evaluation fits Cousin and Whitmore’s (1998) typology as a transformative or emancipatory evaluation approach.

Evaluation for *organizational learning* builds on the foundational work by Argyris and Schön (1978), and later Senge (1990a) that characterizes organizational learning as occurring “when individuals and teams engage in dialogue, reflection, asking questions, and

identifying and challenging values, beliefs, and assumptions (Russ-Eft & Preskill, 2001, p.57). Evaluation researchers such as Cousins and Earl (1995), Preskill and Torres (1999, 2000), Russ-Eft and Preskill (2001) and Torres, Preskill, and Piontek (1996) seek to integrate evaluation into organizational processes to enable continuous improvement and improve an organization's ability to adapt to environmental changes. They position evaluation within the context of an organization and thus emphasize that how evaluation is utilized is mediated by organizational structures and systems. An organization that embeds evaluative inquiry into its internal organizational mechanisms has the ability to be increasingly responsive to the evolving needs of a variety of diverse stakeholders. In practical terms, this positions the evaluator as a facilitator or guide who advocates for evaluation as an ongoing process that is integrated into all work practices and utilized by all stakeholders in a continuous learning process.

Feminist evaluation does not advocate a specific method, but rather outlines an approach toward evaluation that embodies the principles of feminism (Ward, 2002). Feminist evaluation practices acknowledge the political, moral, and ethical underpinnings of evaluative judgment and challenge those facilitating the evaluation process to understand the relationship between voice, perspective, and praxis (Patton, 2002). While feminist evaluation shares many of the goals of other evaluation practices, such as empowerment, utilization-focused, collaborative, and organizational learning, it differs in that it places an unapologetic feminist focus on valuing the voices of women and marginalized groups and explicitly advocates for social change (Beardsley & Miller, 2002; Patton, 2002; Seigart &

Brisolara, 2002; Sielbeck-Bowen, Brisolara, Seigart, Tischler, & Whitmore, 2002a, 2002b; Ward, 2002). Patton (2002) highlights six unique characteristics of feminist evaluation: it explicitly attends to gender inequities and connects this with other forms of social injustice; emphasizes the political nature of evaluation, addressing power relationships throughout the evaluation process; promotes genuine power sharing between evaluatees and evaluator; uses participatory processes to build capacity to sustain process; acknowledges and values women's experiences; uses the evaluation process to facilitate change with a specific focus on social justice (Patton, 2002, p.104).

Summary of the Changing Evaluation Landscape

Reviewing evaluation models over the past four decades reveals significant changes in intended users of evaluation findings, degree of client or stakeholder involvement, underlying assumptions, primary methodology and the major focusing questions (Table 2). As a field, evaluation has evolved from an external top-down purely accountability driven quantitative approach toward a pluralistic approach that incorporates a variety of methods and stakeholder perspectives. While there remains a tension within the field as to how evaluation should be defined and who should be involved, there is a growing consensus that for evaluation to be effective, it is important to include stakeholders throughout the process and critical that evaluation is used by the programs and organization being evaluated. This re-emphasizes that for evaluation to be perceived as useful to an organization and therefore supported, it cannot simply be an accountability mechanism, but should contribute directly to improving the organization.

Changes in the evaluation landscape are reminiscent of a growing awareness in the nonprofit sector of the need to move beyond the emphasis on accountability toward the nonprofit learning from evaluation. The evaluation literature emphasizes that for the nonprofit sector to integrate and use evaluation effectively, nonprofits need to consider the perspectives of multiple stakeholders and not just respond to funding accountability requirements. As a unique organizational context, this challenges the nonprofit sector to discover ways to build its internal capacity to conduct evaluation effectively. This requires an understanding of the link between evaluation and organizational learning so as to understand how evaluation can be an ongoing tool used by nonprofits to self-assess and improve.

The Link between Evaluation and Organizational Learning

As the field of evaluation was moving toward a more complex and pluralistic approach to evaluation that recognized the importance of context and stakeholders to successful evaluation processes, there was a simultaneous push in the corporate sector to use evaluation mechanisms to *improve* rather than to *prove*, which was characterized as *organizational learning*. While the theory of organizational learning has primarily evolved in the for-profit sector, it is useful to explore the history and theoretical underpinnings of the concept of organizational learning in order to better understand how this can be adapted to facilitate the move beyond accountability toward using evaluation for organizational learning in the nonprofit sector. Consequently, a brief review of the theoretical frameworks for

organizational learning will be offered, in addition to a clearer articulation on the linkages between evaluation and organizational learning.

Theoretical Frameworks for Organizational Learning

Organizational learning³ is a complex, multidimensional construct that embraces a systemic approach toward organizations and integrates aspects of several different theories. Webber (1999) traces the chronology of organizational learning from John Dewey's work on experiential learning in 1938, through the introduction of systems thinking to cross-disciplinary audiences at The Macy Conferences in the 1940s, to Kenneth Craik's "mental models," to Lewin's "creative tension" work in 1946, to Jay Forrester's concepts of industrial dynamics in 1961, to Argyris & Schön's (1978) early work around double-loop learning.

These interlinked concepts led to the formation of The Center for Organizational Learning at MIT in 1989, where Peter Senge (one of Forrester's students) served as the Director and published *The Fifth Discipline* in 1990. Senge (1990a) builds on the early work of Dewey, Craik, Lewin, Forrester, and Argyris & Schön, to identify a complex systems approach that emphasizes the importance of feedback loops and incorporates David Bohm's (1989) concepts of dialogue and team learning. Senge (1990a) defines a learning organization as one where "people continuously expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free and where people are continuously learning how to learn together" (p.3).

³ *Organizational learning* (OL) is here used as a construct or process and an organization that embodies the outlined OL characteristics is referred to as a *learning organization*.

This somewhat abstract concept has been increasingly operationalized by more recent scholars from multiple disciplines who have attempted to create frameworks for action that outline the underlying characteristics of organizational learning. Garvin (1993) describes a learning organization generically as an “organization skilled at creating, acquiring, and transferring knowledge, and at modifying its behavior to reflect new knowledge and insights” (p.80). Birleson (1998) refers to a learning organization as an organization that embraces innovation and creates a supportive environment in which learning, adaptation, and continuous improvement can occur. Torres, Preskill, & Piontek (1996) combine these perspectives to define organizational learning as “a continuous process of organizational growth and improvement that (a) is integrated with work activities; (b) invokes the alignment of values, attitudes, and perceptions among organizational members; and (c) uses information or feedback about both processes and outcomes to make changes”(p2). Furthering developing this framework for action, Watkins & Marsick (1996) review 22 case studies of organizational learning and identify eight shared characteristics of learning organizations:

1. Learning organizations focus on organizational learning and transformation; it is not enough for individuals to learn.
2. Structures and systems are created to ensure that knowledge is captured and shared for use in organization’s memory.
3. Leaders and employees at all levels think systematically about the impact of their decisions and work within the total system.
4. Learning is built into work structures, policies, and practices.

5. Learning is transformative in some way, although it is unlikely that some new learning will also be adaptive.
6. Learning has a greater impact when it involves a greater percentage of the employee population.
7. Organizational systems and policies are structured to support, facilitate, and reward learning for individuals, teams, and the organization.
8. Measurement systems benchmark current knowledge and culture and monitor progress toward becoming a learning organization.

Collectively, these frameworks describe the importance of organizational leadership fostering an environment that is conducive to learning at all levels of the organization, integrating this into the systems and practices of an organization, engaging stakeholders and ensuring that their organization is skilled in five main activities: systematic problem solving, experimentation with new approaches, learning from their own experience and past history, learning from the experiences and best practices of others and transferring knowledge quickly and efficiently throughout the organization (Garvin, 1993). Again, while these descriptive and prescriptive frameworks are useful conceptually, they still offer little in terms of the steps toward creating a learning organization and how this intersects with evaluation to impact an organization's performance.

Linking Organizational Learning and Evaluation

Traditional perspectives on evaluation characterize evaluation as a periodic process that involves gathering predetermined information to produce a program, practice, or project

report card that proves success or failure (Gray, 1998). Gray (1998) noted that in 1988, John Brandl, a Minnesota state legislator, gave the keynote address to the American Evaluation Association (AEA) and challenged evaluators to move beyond a focus on projects and programs toward assessing the effectiveness of entire organizations. This initial challenge can be linked to a growing understanding of the importance of making evaluation increasingly participatory, internal, and improvement-focused (Preskill & Caracelli, 1997).

In a review of the past, present, and future of organizational learning and evaluation, Torres and Preskill (2001) traced the focus on organizational learning as a result of the concern from stakeholders about the ways and extent to which evaluation results are used. Torres and Preskill (2001) noted that this concern led to researchers articulating factors influencing use such as timeliness of reports, relevance to stakeholders, evaluator credibility and user characteristics, such as perceptions of the evaluation process (Alkin, Daillak, & White, 1979; Chelimsky, 1986; Cousins & Leithwood, 1986; Newman, Brown, & Rivers, 1983; Patton, 1978; Weiss, 1981). Torres and Preskill (2001) added that while this research was helpful, it did not significantly impact the variable of interest, namely increasing use of evaluation results. Consequently, the evaluation field saw an impetus toward collaborative and participatory evaluation practices in order to increase buy-in, understanding and ultimately use of evaluation findings by the organizations themselves (Cousins & Earl, 1992, 1995; Fetterman, 1996; Greene, 1987; O'Sullivan & O'Sullivan, 1998; Patton, 1986, 1988). While these more collaborative approaches toward evaluation improved evaluation practice, this did not necessarily account for the structural factors that impacted use of evaluation findings by the organization and, as Torres and Preskill noted, this did not position evaluation

as an ongoing tool that could be integrated into an organization and facilitate learning and change at all levels of the organization.

Evaluation professionals such as Cousins & Earl (1995), Patton (1997) and Preskill & Torres (1999) perceived evaluation as a catalyst for organizational learning and emphasized that the evaluation process should be constructed and acted upon by multiple stakeholders, resulting in the findings being used throughout the organization to foster continuous learning. This reconceptualized evaluation as an ongoing, integral—rather than peripheral—part of an organization’s culture, that impacted decision-making at all levels of the organization and was used to help the organization learn and thus improve, rather than “prove,” its operations.

Preskill & Torres (1999) more clearly defined the link between evaluation and organizational learning through their work on evaluative inquiry. Similar to Watkins & Marsick’s (1996) case study review that identified the shared characteristics of learning organizations, Preskill & Torres (1999) identified the elements of an evaluative inquiry approach to organizational learning as including:

- A focus on program and organizational processes as well as outcomes.
- Shared individual, team and organizational learning.
- Education and training of organizational practitioners in inquiry skills.
- Modeling the behaviors of collaboration, cooperation, and participation.
- Establishing linkages between learning and performance.
- Searching for ways to create greater understanding of the variables that affect organizational success and failure.

- Using a diversity of perspectives to develop understanding about organizational issues.

Preskill and Torres (1999) emphasized that this evaluation approach then becomes one that is strength, not deficit based, and becomes a process and a method for asking questions designed to strengthen a organization's capacity for learning and creativity. This makes evaluation a tool for organizational learning as, in addition to answering questions about "how we're doing" and "what works" (or does not) and the impact of this work, it also examines the factors within the organization that enable improved performance, allowing the organization to make changes that will foster results more aligned with mission and related objectives (Merget & Weaver, 1998). Simultaneously, as Leviton (2003) and Botcheva et al. (2002) noted, those involved in the evaluation process "acknowledge that a better understanding of the organizational context and culture in which evaluation takes place is a crucial factor for their success" (p. 422). This again reinforces the importance of engaging staff and multiple stakeholders in the evaluation process as it results in a greater likelihood that the evaluation will meet the needs of the organization and thus be more useful.

Summary of the Theory around Organizational Learning and the Linkages with Evaluation

A learning organization is generically described as an organization that is "skilled at creating, acquiring, and transferring knowledge, and at modifying its behavior to reflect new knowledge and insights" (Garvin, 1993, p.80). Theoretical frameworks by Senge (1990a, 1990b), Garvin (1993), Birleson (1998), Torres, Preskill, & Piontek (1996) and Watkins & Marsick (1996) collectively characterized organizational learning as occurring when there is

supportive leadership fostering an environment conducive to learning at all levels of the organization and ensuring this is manifested in the systems and practices of the organization. This involves engaging multiple stakeholders and ensuring that staff could systematically address problems, experiment with innovative approaches, learn from their own experience, from others, and from organizational past history, and ensure this knowledge is effectively disseminated throughout the organization.

Researchers such as Botcheva et al. (2002), Leviton (2003), Merget & Weaver (1998), and Preskill and Torres (1999) noted that evaluation has increasingly been seen as a useful tool that can facilitate organizational learning. Organizations that can use evaluation successfully become adaptive or innovative as they can clearly articulate inputs, activities, outputs, and outcomes, and use these findings to adapt, improve, and become more effective. However, as Cousins and Earl (1995), Patton (1997), and Preskill and Torres (1999) further noted, this includes a recognition that putting an evaluation system in place is insufficient; there needs to be an organizational infrastructure—in the form of culture, systems and structures, leadership, and communication mechanisms—to support this evaluation process.

Understanding the linkages between organizational learning and evaluation has some significant implications for the nonprofit sector. If nonprofits wish to become learning organizations, then we need to better understand the factors that facilitate nonprofits using evaluation to benefit their organizations. In an article aptly titled “Accountability myopia,” Ebrahim (2005a) commented that evaluation contributes to organizational development (or learning) only when evaluation results in changes in organizational practices; identifying these shortfalls in performance or practice is not enough. This furthermore stresses the need

for moving evaluation beyond simply an external accountability requirement toward a practice that is integrated into the organization and contributes to program and organizational effectiveness. Accountability, which is concerned primarily with effectiveness and how to achieve existing goals and objectives, can be characterized as single loop learning whereas Argyris & Schön (1996) characterize double loop learning as an iterative evaluation process through which organizational values and norms are modified and impact decision-making at multiple levels of the organization. What needs to be better understood is how to approach the central issue of nonprofits both implementing and then using evaluation and integrate these overlapping literatures to comprehend how evaluation can be used as a tool for organizational learning in the context of the nonprofit sector. One way this could be approached is by applying an innovation lens and positioning evaluation as a new tool to be successfully adopted and implemented into the nonprofit sector.

Using an Innovation Lens to Clarify How Evaluation can be Implemented and Used in the Nonprofit Sector

In 1988, Carol Weiss gave a plenary address to the American Evaluation Association (AEA) that made a critical distinction between *doing* and *using* evaluation. Weiss (1988a) emphasized that even when evaluation is sound, credible to program staff, and offers great insights into how to improve programs, there are other barriers—at the individual, group, organizational, and systems level—that may inhibit the use of findings. If evaluation is to contribute to organizational learning in the nonprofit sector, we need to understand the

factors, particularly the organizational factors, which facilitate the doing, or implementation *and* use of evaluation by the program(s) and organization. The innovation literature provides a broad theoretical and empirical framework for thinking about the factors involved in creating organizational change. Using an innovation framework allows us to take a more ecological approach to the doing (or implementing) and using of evaluation and systematically combine the research from multiple diverse fields in order to better understand overlapping factors that may facilitate this change toward integrating evaluation into nonprofits and using evaluation for organizational learning.

Description of the Innovation Process

Rogers (1971, 1983, 1995, 2003) defines innovation as “an idea, practice, or object that is perceived as new by an individual or other unit of adoption.” Building on the notion of stages conceptualized in the early work of Ryan & Gross (1943) on the adoption of new seed corn, Rogers (1983) and other diffusion scholars (Fals Borda, 1960; Fliegel & Kivlin, 1966; Greenberg, 1964; Kincaid, 2000) perceive the innovation decision as a process that occurs over time and consists of five stages: knowledge, persuasion, decision, implementation, and confirmation (Figure 3). To paraphrase Rogers (2003), knowledge occurs upon exposure to and understanding of the function of an innovation, persuasion upon forming a positive or negative attitude toward the innovation, decision upon adopting or rejecting the innovation, implementation upon using the innovation, and confirmation upon reinforcement or reversal of the innovation-decision previously made (pp.169-190).

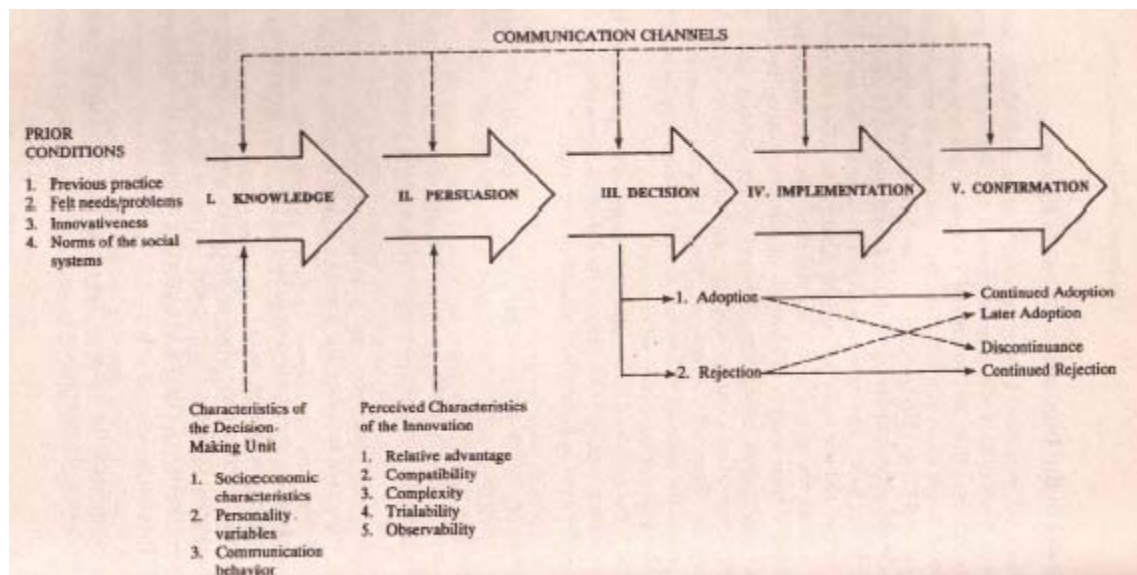


Figure 3: A model of the five stages in the innovation-decision process (Rogers, 2003, p.170).

Empirical support for the validity of this stage model comes from Beal, Rogers, & Bohlen (1957), Coleman, Katz, & Menzel (1966) and Kohl (1966). Participants in these studies all reported passing through stages during the innovation-decision process, though Rogers (1983) acknowledges that while the evidence is clear for the knowledge and decision stages, there is limited data on the distinctiveness of the implementation and confirmation stages (p.198). Nonetheless, similar stages models have been utilized in the preventive health field, including Prochaska's (1979) stages-of-change model—moving through *precontemplation*, *contemplation*, *preparation*, *action*, and *maintenance*—and Weinstein and Sandman's (2001) *precaution adoption process* model. Again, while there is some support for individuals moving through stages during the innovation process, characterized by Prochaska (1979) and Weinstein and Sandman (2001) as health behavior change, limited

evidence again exists for the distinctiveness of these stages and there is an overall lack of understanding of how these stages interrelate.

More importantly, a further criticism of stage models comes from Tornatzky & Fleischer (1990), who draw on a body of research emanating from Fairweather's theory of social change (1972) and Experimental Social Innovation and Dissemination (ESID) model (1967) to critique the linear, individual-level analysis; offering instead a more ecological perspective of innovation. Tornatzky and Fleischer's focus is on technological innovation, which involves the "situationally new development and introduction of knowledge-derived tools, artifacts, and devices by which people extend and interact with their environment" (Tornatzky & Fleischer, 1990, p.11). This emphasizes that technological innovation impacts and is impacted by the surrounding socio-cultural-environmental system into which it is implemented. While much of the literature surrounding technological innovation uses the context of for-profit organizations, many of the principles can be applied and measured in a nonprofit context. Akin to the structure of for-profits, nonprofits are embedded in a complex interconnected system that both impacts and is impacted by organizational activities and interactions at the individual, group, organizational and societal level, both internal and external to the nonprofit (Bamberger, 1991; Hughes, 1994; Kline, 1986; Tornatzky & Fleischer, 1990).

Tornatzky & Fleischer (1990) do not negate that there are stages or more appropriately a *life cycle* of innovation, but emphasize the importance of recognizing that innovation is a "complex interaction of people, scientific concepts, aspirations, and consequences...understood at many different levels of aggregation" (pp.49-50). In line with

previous work (Tornatzky, Eveland, Boylan, Hetzner, Johnson, Roitman, and Schneider, 1983) Tornatzky & Fleischer differentiate between *source-centered* and *user-centered* stage models, describing source-centered models as related to and from the perspective of developing technology and user-centered models as reflective of the perspectives of the user, such as Rogers' innovation-decision model. Thus source-centered models focus on technology creation and dissemination, while user-centered models focus on systemic integration of this technology through adoption, implementation, and routinization. In the larger context of the innovation process, these spheres overlap and Tornatzky & Fleischer (1990) reinforce that these are not isolated chronological concepts and, as figure 4 shows, should practically be viewed at multiple levels of analysis as interconnected parts with “multiple feedback and feedforward cycles of information exchange” (p.30).

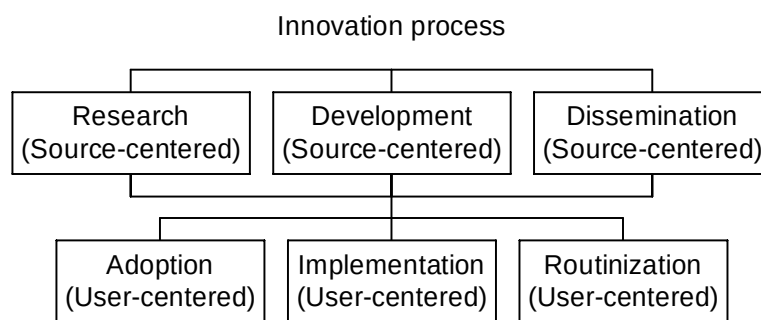


Figure 4: Interconnected multi-leveled processes of technological innovation (Tornatzky & Fleischer, 1990).

A more useful conceptual model for the technological innovation process would thus extend (and feed back) from research and development through technology adoption and implementation, product and service design, investment and employment policies, and global

and local activities, as well as to issues of management and government relations (Brown, Green, Hall, Rocchi, Rutter, & Dearing, 2000). While Kline's (1986) chain-linked model emanates from the perspective of the developer, the complexity and interrelation of this innovation process are demonstrated through the prominent feedback loops (Figure 5).

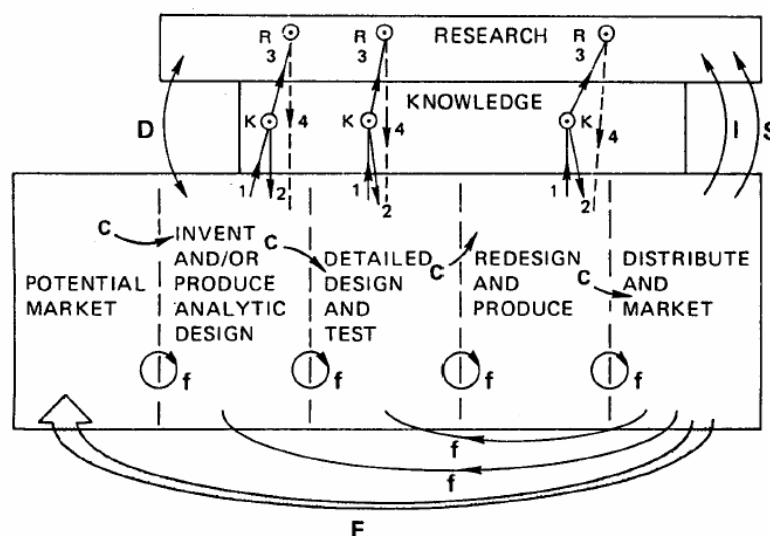


FIGURE 3 Chain-linked model showing flow paths of information and cooperation. Symbols on arrows: C = central-chain-of-innovation; f = feedback loops; F = particularly important feedback.

K-R: Links through knowledge to research and return paths. If problem solved at node K, link 3 to R not activated. Return from research (link 4) is problematic—therefore dashed line.

D: Direct link to and from research from problems in invention and design.

I: Support of scientific research by instruments, machines, tools, and procedures of technology.

S: Support of research in sciences underlying product area to gain information directly and by monitoring outside work. The information obtained may apply anywhere along the chain.

Figure 5: Chain-linked model showing paths of information and cooperation (Kline, 1986).

For the purposes of looking at evaluation in the context of the nonprofit sector, what is vital to distill from this model and other dynamic models (e.g. Havelock, 1973) are the multiple feedback loops these models offer, which reinforce the complexity and non-linearity

of the innovation paradigm. In addition, as Tornatzky & Fleischer (1990) continually emphasize, this complexity is compounded by interactions at the individual, group, organizational and societal level, both internal and external to the unit of analysis; combined this creates the social context for *technological* innovation. Critical to the development of an innovation model that seeks to integrate a *social* innovation, such as evaluation, into nonprofits so as to increase use by the nonprofits themselves is an understanding of how principles from these interdependent source- and user-centered components could be used as a framework to better understand and plan for the innovation process.

Social Innovation Model

Mayer and Davidson's (2000) model of social innovation provides a clearer framework in which to understand the factors that may facilitate nonprofits using evaluation for organizational learning. Mayer and Davidson (2000) draw on the work of Fairweather and colleagues (Fairweather, 1967, 1972; Fairweather & Tornatzky, 1977; Fairweather, Sanders & Tornatzky, 1974; Fairweather & Davidson, 1986; Rappaport, 1977; Tornatzky, Fergus, Avellar, Fairweather & Fleischer, 1980) to apply an innovation lens to look at social programs and outcomes as innovation and dynamically link the three interdependent components of implementation, adoption and routinization. In a social program context, adoption involves the host setting's decision to use a new program, implementation refers to the actual use of that program, and routinization to the seamless integration of this program into the setting (p.4). Mayer & Davidson's (2000) model further recognizes that there is no

automatic link between implementation, adoption and routinization, creating a need to better understand these processes.

By positioning evaluation as a social innovation that takes place in the context of a nonprofit organization, we are challenged by Tornatzky and Fleischer (1990) to see the implementation of evaluation as a process that impacts and is impacted by the surrounding socio-cultural-environmental system into which it is implemented. Fixsen et al. (2005) cite evidence from Bauman, Stein, and Ireys (1991), Dale, Baker, and Racine (2002), and Winter and Sulanski (2001) that the more clearly the core components of a social innovation can be defined, the more readily the program or practice can be implemented successfully (p.24). This supports the research of Mayer and Davidson (2000) and provides further insights into ways to differentiate between doing or implementing evaluation and using evaluation. While the initial stage of adopting and implementing evaluation involves simply *doing* evaluation, a core component of evaluation is that it involves the *use* of evaluation by organizations beyond accountability so as to contribute to organizational learning. Whether an organization benefits from evaluation, rather than simply using it in a compliant way as a tool for accountability, may depend on the extent to which it is using evaluation internally. Thus, only when evaluation is used by the organization to improve programs and practices is there a degree of *fidelity* to the implementation of evaluation for organizational learning.

Fidelity and Core Characteristics of Evaluation Implementation and Use

The debate surrounding the importance of fidelity (the degree to which the innovation is implemented in a way that replicates the original model) versus reinvention

(the degree to which an innovation is modified or changed during the process of adoption and implementation) ranges from research suggesting the need for fidelity to the original model (Caslyn, Tornatzky, & Dittmar, 1977; Hall & Loucks, 1977) to a belief in the need for reinvention (Emrick, Peterson, & Agarwal-Rogers, 1977; Larsen & Agarwal-Rogers, 1977).

One of the most comprehensive studies on this subject conducted at the Center for Innovation Research at the University of Michigan using a multi-innovation, multi-organizational sample indicated that replicate studies that went beyond the original model yet retained a high fidelity to that model were more likely to achieve positive outcomes than those that modified the model (Mayer & Davidson, 2000). Blakely, Emshoff, & Roitman (1984) reinforced this finding by suggesting that fidelity to the model was important, but reinvention in the form of additions could improve upon the model and have a positive effect on outcomes. Rogers (2003) largely ignores these findings and cites research that indicates that a higher degree of re-invention leads to a faster rate of adoption (Backer, 2000) and a higher degree of sustainability (Goodman, Steckler, & Kegler, 1996; Rogers, Peterson, Cunningham-Sabo, & Davis, 2004).

While this fidelity-reinvention debate continues, some consensus is being built around understanding the *core elements* of an innovation in terms of the features that are responsible for its effectiveness (Kelly et al., 2000; Rogers, 2003; Tornatzky & Fleischer, 1990). By emphasizing that use of evaluation by programs and nonprofit organizations is a core component of fidelity to the innovation (evaluation for organizational learning), the current research is clarifying that *internal use* is essential to achieving the outcomes desired by nonprofits (Fixsen et al., 2005).

Another core component of using evaluation for organizational learning is the inclusion of diverse stakeholder perspectives throughout the evaluation process, which could range from participation in the process to complete ownership of the evaluation process by the nonprofit staff. The importance of stakeholder engagement in the innovation process re-echoes the values underpinning any evaluation that is focused on use by organization members, such as utilization-focused (Patton, 1978), participatory or collaborative (Cousins & Earl, 1992, 1995), empowerment (Fetterman, 1994), feminist (Ward, 2002), and organizational learning (Preskill & Torres, 1999). All of these evaluation models emphasize that program people are a key part of implementation and use of evaluation and that their inclusion throughout the process increases the likelihood of the organization successfully implementing evaluation and learning from and using the evaluation findings. In this way, stakeholder integration can be seen as an essential component of evaluation for organizational learning that is necessary to enable successful implementation and internal use of evaluation by the nonprofit.

Factors that Impact the Innovation Process

Perceiving evaluation as a social innovation helps to clarify the distinction between implementation for accountability and implementation leading to use by nonprofits. Positioning use of evaluation by nonprofits as a core component of implementation that demonstrates fidelity to this new model of evaluation for organizational learning allows us to articulate the factors that are highlighted throughout the innovation literature as important to the innovation process. This requires looking at the relationships between the different

stages of the innovation process and assessing what impacts each stage and the overall innovation process.

Tornatzky & Fleischer (1990) characterize implementation processes as including identifying core aspects of the innovation (such as use of evaluation by the nonprofit), measuring to what degree these core aspects are implemented, and distinguishing between the positive and negative aspects of adaptation of an innovation. For the purposes of looking at the implementation of evaluation in a nonprofit context, if evaluation is mandated externally by the government or other funding group, then implementation may well be restricted to accountability purposes with the nonprofit simply doing evaluation for external purposes or cooperating with an external evaluator rather than routinizing it seamlessly into organizational practice. In contrast, if a nonprofit instead adopts evaluation as it perceives the myriad ways it can help the nonprofit develop and learn, then implementation with a high degree of fidelity may well result in the use of evaluation for organizational learning. What is therefore important to understand from the literature are the innovation characteristics that are deemed important for the success of implementation. These will help to understand how to increase the chances of a successful routinization process (Bartholomew, Parcel, Kok, & Gottlieb, 2001; Howze & Redman, 1992; Oldenburg & Parcel, 2002).

Innovation Characteristics for Implementation

Research on innovation characteristics accentuates the importance of ensuring both the social innovation (evaluation) and the way it is implemented are compatible or congruent

with the needs, values, and practices of the organization, that the advantages are communicated in a way that is relevant to the organization, using appropriate communication networks and that the social innovation is presented in a clear and understandable way that demonstrates its ease of use (Klein & Sorra, 1996; Rogers, 1995, 2003; Tornatzky & Klein, 1982).

In a nonprofit organization, this restates the importance of understanding the unique tensions within a nonprofit context that is focused on providing services and working with multiple internal stakeholders to emphasize the program and organizational advantages of using evaluation. In addition, one factor deemed critical to the implementation of evaluation within the context of the nonprofit sector is the importance of gatekeepers at multiple levels of the organization. These include linkage agents who connect those promoting to those using the innovation, such as program heads (Monahan & Scheirer, 1988; Goodman, Steckler, & Kegler, 1996) and existing gatekeepers for information delivery systems, such as nonprofit executive directors and other leaders, program staff, the nonprofit board, and other stakeholders, such as funders, who may add credibility to the importance of evaluation in the nonprofit (Oldenburg & Parcel, 2002).

Implicit in the discussion of gatekeepers and organizational-innovation congruence is the recognition of the need to involve more stakeholders in the process of innovation and to encourage feedback on the perceived or expected negative consequences of the social innovation (Roome, 1998, pp.172-181). As the findings of both Rogers (1971, 1995, 2003) and Tornatzky & Klein (1982) show, early and continued inclusion of stakeholders in the

innovation process may improve implementation strategies, and encourage social acceptance.

Nadler (1983) stresses the need to recognize that if organizations are context-dependent systems, then including stakeholders in the design of the change can assist the implementation process, reducing any resistance by shaping the political dynamics of change so that power centers develop that support, rather than impede, change. Klein and Sorra (1996) further reinforce the importance of involving internal stakeholders early on in the implementation phase as it can improve employees' shared perceptions of events, practices, procedures and behaviors that are rewarded, supported and expected, thus promoting a climate for effective implementation. This emphasizes that using multiple viewpoints at multiple levels can clearly impact the innovation process at all stages and that the earlier these perspectives are integrated, the more effective they will be.

Summary of the Innovation Process as it Relates to Evaluation

For the purpose of looking at evaluation as a social innovation in a nonprofit organization, what is critical to distill from this literature is that doing and using evaluation are interdependent, but still different components of the evaluation for organizational learning process. While the “doing” of evaluation could simply mean cooperating with external evaluators and being passively accountable, the active use of evaluation by the nonprofit is a measure of fidelity to implementing this type of evaluation. If nonprofits voluntarily make the decision to implement evaluation activities based on the perceived advantages gained from evaluation, they may be more likely to use evaluation for

organizational learning. In this type of evaluation, characterized as evaluation for organizational learning, internal use of evaluation by the nonprofit is positioned as a core component of implementing the innovation (evaluation for organizational learning) into the nonprofit sector.

Successful implementation of evaluation that results in use of evaluation to inform organizational practice requires a stronger understanding of the social structures and processes and the prevailing socio-cultural-economic-environmental paradigm in order to better identify the core characteristics of both the innovation of evaluation and the nonprofit context that impact positive outcomes. This emphasizes the importance of understanding the unique elements of the nonprofit organizational context in order to comprehend what factors may impact how effectively evaluation is integrated and used by the nonprofit.

The innovation literature highlights three crucial factors that could potentially impact implementation of evaluation for organizational learning in the nonprofit sector: Gatekeepers and/or champions, such as those in positions of leadership, can play key roles in developing organizational culture and establishing evaluation as a priority; early and sustained inclusion of multiple and diverse stakeholders is critical throughout the evaluation process; and it is crucial to understand that the evaluation process impacts and is impacted by the complex interconnected system in which the nonprofit is embedded in addition to characteristics of the organization itself, such as size, age, function, and whether a culture of learning exists.

Summary: Toward a Conceptual Framework for Evaluation for Organizational Learning in the Nonprofit Sector

If we approach evaluation through an innovation paradigm, we can start to identify factors that may facilitate nonprofit organizations using evaluation for organizational learning. If evaluation for organizational learning rather than accountability is positioned as a social innovation, then successful implementation into an organization requires a stronger understanding of the socio-cultural-economic-environmental structures and processes so as to understand both the core characteristics of the social innovation (evaluation for organizational learning) and the context (the nonprofit organization) that impact the evaluation process. This innovation lens further emphasizes the importance of gatekeepers, such as leaders, stakeholder participation, and organizational culture or values in promoting the use of evaluation for organizational learning. This necessitates an understanding that the process of integrating evaluation into an organization is complex, operating at multiple levels, and both impacts and is impacted by the activities at the level of the individual, group, organization, and larger social environment (Bamberger, 1991; Havelock, 1975; House, 1981; Hughes, 1994; Kline, 1986; Tornatzky & Fleischer, 1990).

Congruent with Klein and Sorra's (1996) notion of the importance of innovation-values fit, the literature around the nonprofit sector context emphasizes the tension between applying limited resources to programs and services and investing in organizational capacity that could facilitate improved use of evaluation by the nonprofit. Nonprofits have distinct structural and legal characteristics and are organizations that are mission driven, usually toward social goals. Even with the external pressure for accountability, this social focus

potentially makes the sector resistant to practices such as evaluation, as they detract from the nonprofit's mission of providing much needed services in the community. Therefore, the use of evaluation by the nonprofit is a function of the perceived *fit* of evaluation to organizational and nonprofit staff's values. Looking at the nonprofit sector from a historical perspective helps to characterize the move from accountability toward integrating evaluation into organizations and also identifies tensions that may facilitate or hinder this process. Understanding this unique context is important as it helps identify factors that should be included in any nonprofit research, such as organizational size, age, the role of leadership, geographic service area, field and funding sources.

Looking at the field of evaluation helps to identify core characteristics that show fidelity to the implementation of the social innovation (evaluation for organizational learning) and how it can be integrated into nonprofit practices to aid in program and organizational development. Evaluation has evolved into a pluralistic approach that recognizes that ongoing participation by diverse stakeholders and use of results by the organization and its programs are key components of the evaluation process. If evaluation is to be both used and useful, it must reflect the organizational context and be used to improve programs, practices, and organizations and not simply be a tool that is used for external accountability. This necessitates a distinction between implementation and use of evaluation as *doing* or implementing evaluation does not always equate to *using* the results within a nonprofit. Doing evaluation could simply be a nonprofit complying with external accountability requirements and even simply cooperating with an external evaluator. In contrast, use of evaluation for organizational learning involves the active participation of

multiple stakeholders and the internal use of evaluation by the nonprofit. Consequently, implementation and use of evaluation should be treated as separate variables in any nonprofit research focused on the use of evaluation for organizational learning.

To move toward a model of using evaluation for organizational learning and not just for accountability, nonprofits need to have an infrastructure in place that supports the evaluation process. Learning organizations have been characterized as having supportive leadership that helps build the organization's capacity to engage stakeholders and ensure that staff can systematically address problems, experiment with innovative approaches, learn from individual and organizational experiences, and ensure this knowledge is transferred throughout the organization. What is not known is how all these factors—organizational learning, stakeholder engagement, supportive leadership, and evaluation implementation and use—interrelate and facilitate the integration of evaluation for organizational learning purposes. The next chapter will take a closer look at the research surrounding these constructs and any associations between organizational learning, stakeholder engagement, supportive leadership, and evaluation implementation and use.

III: LITERATURE REVIEW

Introduction

While there is a growing body of empirical literature connected to the nonprofit sector and to evaluation, there is no existing literature that has used a social innovation framework to look at the intersections between supportive leadership, stakeholder engagement, organizational learning, and implementation and use of evaluation in the nonprofit sector. Consequently, this review will draw on the available empirical literature to develop a theoretical framework that brings these factors together in the context of the nonprofit sector in order to elicit a proposed model of these relationships and articulate a set of research questions.

While the empirical literature related to the nonprofit sector is limited, this section will provide an overview of nonprofit and other relevant studies and then outline the importance of further research in this area. This section will begin with definitions of implementation and use of evaluation and provide empirical research on these two evaluation process stages. Research will be offered in support of the dimensions underlying organizational learning and the research links between organizational learning and evaluation will be further addressed, in addition to why stakeholder engagement and the role of supportive leadership should be factors in any work concerning evaluation and organizational learning.

Three main approaches were used to obtain a comprehensive list of literature that blanketed this proposed research area. Initially, I conducted computerized bibliographic searches of a variety of databases, including Business Source Premier, LexisNexis Academic,

Web of Science (Social Sciences Citation Index), Social Science Index, and PsycINFO using a variety of keywords for subject, title, then abstract, usually extended to full text. Each section will highlight the keywords used to elicit the research connected to that area of research. The bibliographies and reference lists of any articles that were found in these areas were then searched for additional literature. Finally, to be as thorough as possible, a variety of nonprofit and evaluation related journals were searched to ensure I had not missed relevant articles. These included, but were not limited to: *American Journal of Community Psychology*, *American Journal of Evaluation*, *Business and Society*, *Evaluation Review*, *New Directions in Evaluation*, *Nonprofit and Voluntary Sector Quarterly*, *Nonprofit Management and Leadership*, *Nonprofit World* and the former *Journal of Volunteer Administration*.

Implementation and Use of Evaluation

To identify literature relevant to the implementation and use of evaluation in relation to nonprofits or other organizations, I used the three approaches as defined in the introduction above using the keywords adoption, implementation and/or use refined by combinations of the keywords adoption, implementation and/or use, evaluation, nonprofit, knowledge utilization, social innovation and dissemination. While it would have been useful to find studies that dealt with samples of nonprofits, only a few studies identified their sample as nonprofits, and in other studies, it was not always made apparent whether organizations that could potentially be nonprofits (such as community agencies) were registered nonprofits. Consequently, while every attempt will be made to distinguish between nonprofit and for-profit studies, this may not always be possible. This section will introduce definitions of implementation and use of evaluation, review theoretical frameworks

around factors that impact evaluation use and offer a critique of available and relevant literature on implementation and use of evaluation.

Definitions of Implementation

Fixsen, Naoom, Blase, Friedman, and Wallace (2005) define implementation as “a specified set of activities designed to put into practice an activity or program of known dimensions” (p.5). Klein and Sorra (1996) define implementation as the “process of gaining targeted organizational members’ appropriate and committed use of an innovation” (p.1055). These definitions emphasize that it is critical to evaluation implementation to know *what* activities are being implemented in addition to *how* they are being implemented and by *whom*. While much of the literature around implementation revolves around the implementation of interventions in the form of programs and policies, this concept can be ascribed to the implementation of evaluation in nonprofit organizations. As Weiss (1998a) and Ebrahim (2005a) noted, there is a significant difference between doing (implementing) and using evaluation results in a way that allows an organization to learn. *Implementation of evaluation practices* is a necessary precursor to the more complex construct of *use of evaluation results* (i.e. if an organization is not doing any evaluation, there are no results to use), but is unknown whether implementation automatically results in use by a nonprofit or other organization. Consequently, it is essential to understand the core characteristics of what is being implemented in order to better comprehend how this interacts with other factors to impact use of evaluation for organizational learning.

Definitions of Use and Theoretical Frameworks around Factors that Impact Evaluation Use

One ongoing critique of the research around evaluation use is the considerable variability in how use is operationalized as a dependent variable (Cousins & Leithwood, 1986; Leviton, 2003). While some studies look at evaluation use instrumentally as support for decisionmakers or as a tool to educate various stakeholders, others more ambiguously measure processing of evaluation results as use, or simply focus on utilization potential. Leviton and Hughes (1981) distinguished between three types of evaluation use: instrumental (for decision-making), conceptual (to educate) and symbolic (for political purposes). Greene (1988a) used qualitative data from two case study evaluations to describe a typology of evaluation uses that divided use of evaluation results in program evaluation (Table 3). Based on documented uses from her two case studies, Greene (1988a) articulated five different major groupings: minor instrumental use, major instrumental use, conceptual use, persuasive use and symbolic use. While there were interconnections between these uses, Greene (1988a) distinguished between them to start to more fully articulate and operationalize the “use” construct. To date, no further research around evaluation utilization has used these definitions to operationalize and measure the concept of use.

Table 3

A Typology of Use of Evaluation Results

Use of Evaluation Results	Example
Major instrumental use	Significant new program development, policy implementation, planning activities
Minor instrumental use	Smaller procedural changes in program operations and activities
Conceptual use	Broader and deeper program understanding, representing important confirmation of existing intuitions
Persuasive use	Citation of results in reports to external audiences
Symbolic use	Enhanced prestige and visibility for the program within the larger community

Note. Adapted from Greene, 1988a, p.100

In a survey of the Evaluation Use Topical Interest Group (TIG) members, Preskill and Caracelli (1997) attempted to ascertain members' perceptions about and experiences with evaluation use. Using a mail survey sent to the 530 Evaluation Use TIG members, Preskill and Caracelli (1997) provided a set of close-ended questions to ascertain views on evaluation, evaluator roles, evaluation use and misuse, and changes in evaluation practice. From 282 respondents (response rate of 54 percent), Preskill and Caracelli (1997) generated descriptive statistics and found that the majority of respondents agreed that the major purposes of evaluation were to provide information for decision making (99 percent), improve programs (99 percent), facilitate organizational learning (88 percent) and improve the knowledge or worth of the evaluand⁴ (79 percent). In terms of strategies that would facilitate use, Preskill and Caracelli (1997) found that critical factors included: planning for

⁴ *Evaluand* is defined as the thing (i.e. program, practice, method, proposal) that is being evaluated

use at the beginning of an evaluation (90 percent), involving multiple stakeholders in the evaluation process (72-95 percent), identifying and prioritizing intended users and intended uses of the evaluation (79-85 percent), designing the evaluation within resource limitations (83 percent), communicating findings to stakeholders (79 percent) and developing a communication and reporting plan (79 percent). Reinforcing the research of Shulha and Cousins (1997), Preskill and Caracelli (1997) proposed from these findings that views of evaluation use seem to have shifted significantly since the early program evaluation research and now include the importance of evaluation context and processes.

In his work on utilization-focused evaluation, Patton (1997) added process use to Leviton and Hughes' (1981) and Greene's (1988a) notion of instrumental use, which Patton differentiated from instrumental use such as judging merit or worth, improving programs and generating knowledge (Russ-Eft, Atwood, & Eggherman, 2002). For Patton (1997), process use is "indicated by individual changes in thinking and behavior, and program or organizational changes in procedures and culture, that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process" (p.90). Again, one of the complications of utilizing *use of evaluation* as a construct is that instrumental use and process use are extremely interrelated and what is difficult to discern is where process use ends and instrumental use begins.

Henry and Mark (2003) and Kirkhart (2000) took this critique of use one step further by highlighting that in addition to being overlaid with multiple meanings and unclear levels of analysis, use is focused on to the detriment of types of influence that occur over time and result in long term program, organization and social changes. Instead, Henry and Mark

(2003) drew from the social science change process literature and offered a framework of causal “influence pathways.” This framework defined levels of influence at the individual, interpersonal, and collective level and then characterized specific types of change processes or outcomes that could plausibly be related to evaluation. While this adds to the multidimensionality of the construct of use, it offers little in terms of concrete operationalization of use as a variable. Despite the repeated critiques by Cousins and Leithwood (1986), Leviton (2003) and Shulha and Cousins (1997), Henry and Mark (2003) offered nothing in the form of empirical data to support their hypothesized constructs and to date, no other work has used these definitions to operationalize and measure the construct of use.

In an attempt to move toward a theoretical model of evaluation utilization, Johnson (1998) integrated literature from multiple fields, analyzed implicit (developed from evaluation writings) and explicit (already published) evaluation utilization process-models, and developed a meta-model of evaluation utilization by multiple organizational stakeholders (figure 6). Johnson (1998) defined implicit process-models as models that do not specify variable ordering and explicit models as process-models that have been empirically tested.

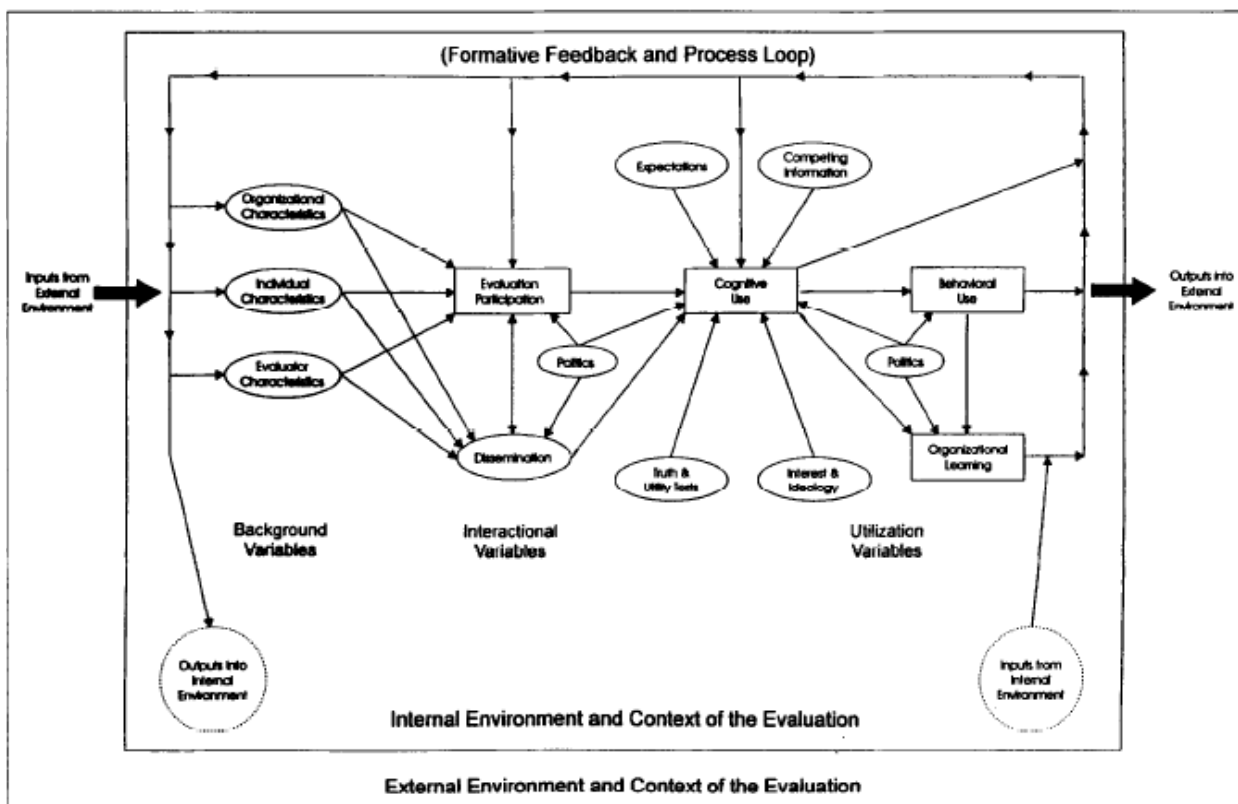


Figure 6: A theoretical model of evaluation utilization (Johnson, 1998, p.104).

Based on these models and ideas from complexity theory, organizational learning and organizational design, Johnson (1998) developed a dynamic model of evaluation utilization that defined evaluation utilization as a “continual and diffuse process that is interdependent with local contextual, organizational and political dimensions. Participation by stakeholders is essential and continual (multi-way) dissemination, communication and feedback of information and results to evaluators and users (during and after a program) help increase use by increasing evaluation relevance, program modification and stakeholder ownership of results. Evaluators, managers, and other key stakeholders should collaboratively employ organizational design and development principles to help increase the amount and quality of

participation, dissemination, utilization and organizational learning” (p.104). Johnson’s (1998) model was useful only in that it more clearly outlined the complexity of the multiple factors, both internal and external to the organization, that impacted cognitive and then behavioral use of evaluation. As with the work by Greene (1988a, 1988b), despite an extensive search through the literature, no empirical work has been conducted to test Johnson’s (1998) propositions and refine these theoretical relationships and Johnson (1998) made no effort to hypothesize which factors may be more or less important in influencing use. Consequently, while Johnson (1998) articulated the complexity of the factors that impact use, there is neither theoretical nor empirical support for the relative importance of each of these factors in influencing use.

Summary of Definitions of Use and Theoretical Frameworks

While there is more than 35 years of research around use of evaluation, there continues to be a lack of consensus around the operationalization of *use* as a construct. While there has been a significant amount of theoretical discussion and critique around the construct of use by Greene (1988a, 1988b), Henry and Mark (2003), Johnson (1998), Kirkhart (2000) and Patton (1997), this has not resulted in empirical work to substantiate their theoretical premises. Consequently, there is some consensus at the global level that evaluation use is multidimensional and engagement in this process can impact individuals, groups, and organizations, leading to myriad cognitive and behavioral uses. In addition, there is a growing understanding that evaluation use is significantly impacted by evaluation context and processes. However, a lack of empirical data surrounding this construct make it

less clear *who* predominantly uses evaluation, *how* evaluation is most typically used, and *which factors* facilitate that process.

For a nonprofit to move beyond simply implementing evaluation and being accountable to funders toward using evaluation for organizational learning, who is involved in the implementation process and how evaluation information is used (and by whom) is critical. Evaluation that is used simply as an accountability tool with results being either shelved or reported externally to funders or the board is different from evaluation that feeds back into an organization and is seen as a useful tool by multiple stakeholders, both internal and external to the nonprofit, that can be used to improve programs and organizational practices. Consequently, while there is clearly some ambiguity around the construct of use, in the following review every attempt will be made to clearly articulate how the authors are stating *who* is using the evaluation results (the program/organization or funders and other external stakeholders) in addition to *how* use of evaluation results is being defined (as an accountability mechanism or as a way for organizations to improve programs and practices).

Empirical Work on Evaluation Implementation

In a qualitative study based on semi-structured interviews with eighteen university outreach program staff engaged in one program, Taut and Alkin (2003) questioned whether factors that act as a barrier to evaluation use by programs are also barriers to evaluation implementation when an external evaluation is conducted. Given that implementation prefaces use, it is not surprising that they found congruence between factors impacting implementation and utilization. Akin to the work of Klein and Sorra (1996), Taut and Alkin

(2003) highlighted several factors that need to be considered in any research looking at barriers to evaluation implementation, which mainly revolve around the failure to include stakeholders in the evaluation process resulting in a disconnect between the evaluator and the program staff leading to low user belief and lack of trust in the evaluation process and findings. While this study is hampered by a low sample size and information around a single program, the insights are congruent with the literature around the importance of stakeholder engagement *throughout* the evaluation process. If program staff that are going to use evaluation are not included throughout the evaluation process, there is less likelihood that the evaluation results will be perceived as credible or even that they will be asking the questions deemed important to program staff.

In a series of ongoing mixed method studies, Carman (2005) and Carman and Millesen (2004) used data gathered from interviews with executives from nonprofits and a mail survey sent to a random sample of nonprofits from the fields of community development, developmental disabilities, and social services in the states of New York and Ohio to explore what nonprofits were doing in regards to evaluation, why they engaged in those efforts and how they used the results. Using a comparative framework to look at differences between these types of organizations, Carman (2005) and Carman and Millesen (2004) noted that there was significant variability in how evaluation was both defined and operationalized across the nonprofit sector. During interviews conducted with 31 nonprofit executives within New York State, Carman (2005) noted that nonprofit executives thought very broadly about program evaluation and performance measurement activities, including activities that were related to reporting, monitoring, management, and government

regulations. Consequently, Carman (2005) highlighted the role that these executive directors played in this process, finding that content analysis of qualitative data suggested that implementation of evaluation practices was related to the perceptions of the nonprofit executives themselves, whether they had experience with evaluation and/or a positive perception of what it could contribute to their organization.

Carman (2005) triangulated her qualitative analysis with quantitative data from 178 respondents representing the same three nonprofit social service areas: community development, developmental disabilities and social services. This quantitative data seemed to further emphasize differences between which types of organizations were gathering “higher level” evaluation data such as outcomes and consumer satisfaction data. Carman (2005) noted that 80 percent of organizations providing services to people with developmental disabilities reported that they gathered outcome data compared to 60 percent of the social service and community development organizations. This difference was replicated with consumer satisfaction data, as 95 percent of organizations providing services to people with developmental disabilities reported that they gathered consumer satisfaction data compared to 66 percent of the social service and 53 percent of the community development organizations. Consequently, it will be critical to assess which evaluation activities are implemented within nonprofits and how this differs by type of nonprofit.

Summary of the Theory and Research around Implementation of Evaluation

The theoretical and empirical research around implementation of evaluation suggests that it is critical to understand core aspects of the evaluation process. This includes what

evaluation activities are being implemented in addition to how widely evaluation is being implemented across the organization, in terms of the prevalence across organizational practices and programs. Qualitative research by Taut and Alkin (2003) emphasized the importance of stakeholder engagement throughout the evaluation process as a way to increase implementation (and use by program staff). Carman (2005) and Carman and Millesen (2004) used a mixed method approach to explore comparative evaluation practices and found that both the qualitative and quantitative data showed there was considerable variability in how evaluation was being both defined and implemented across nonprofits. Based on interviews conducted with 31 nonprofit executives, Carman (2005) highlighted the role that these executive directors played in this process, finding that qualitative data showed that if executive directors had experience with and positive perceptions of the benefits of evaluation, they were more likely to support the implementation of evaluation. This re-emphasizes the importance of including executive directors' attitudes toward evaluation as it could impact implementation of evaluation in nonprofits. What is not known or addressed in this research on the nonprofit sector is how evaluation practices can be parsed out to create a scale of implementation so as to look at how differential implementation of evaluation activities impacts use by nonprofit program staff. Consequently, it is unknown how implementation relates to use, which factors impact implementation and how this relationship is associated with a nonprofit using evaluation results for organizational learning.

Empirical Literature on Evaluation Use: 1971-1996

Leviton and Hughes (1981), Cousins and Leithwood (1986) and Shulha and Cousins (1997) conducted three extensive syntheses on evaluation use. As these syntheses collectively review over 120 studies related to evaluation use from 1971 to 1996, it should be sufficient to document the collective findings from these studies in order to focus on more recent empirical work.

Leviton and Hughes (1981) conducted a critique on the empirical literature surrounding the utilization of evaluations by policymakers and organization leaders predominantly in the governmental and for-profit sectors, paying particular attention to methodological issues around the definition of utilization and methods for detecting utilization. Leviton and Hughes (1981) criticized the lack of empirical work around evaluation utilization, noting that much of this work on utilization is limited to individual case studies or on policymakers' statements in interviews and surveys, leading to a lack of generality and confounding of situational factors. In addition, Leviton and Hughes (1981) highlighted the biases associated with retrospective research, as interviewees may remember dramatic instances of use more easily than frequent, but modest ones. The major methodological issue to Leviton and Hughes (1981) was the lack of agreement around the operationalization of "utilization" as a unit of analysis.

Based on their review of the empirical literature and with the disclaimer of the weakness of the utilization construct, Leviton and Hughes (1981) concluded there were five clusters of variables impacting utilization: the relevance of evaluation to the needs of potential users, the extent of communication between potential users (such as program staff)

and producers (such as external evaluators) of evaluations, the translation of evaluations into their implications for policy and programs, credibility or trust placed in evaluations and commitment or advocacy by individual users. Reminiscent of Roger's (1983, 1995, 2003) diffusion work and Tornatzky and Fleischer's (1990) dissemination research, Leviton and Hughes emphasized that evaluation would be more used if evaluation was relevant (addressed client's needs in a timely fashion), disseminated widely with clearly presented information that stakeholders can comprehend and therefore use, credible and conducted in an environment where there is a commitment to evaluation and advocacy for use by key individuals. What was problematic about Leviton and Hughes (1981) critique was that the authors did not make evident how these clusters of variables impacting utilization were relevant to *who* the users were and whether it was important whether the producers of evaluation were internal or external to the organization.

To describe research-based knowledge surrounding evaluation use, Cousins and Leithwood (1986) meta-analyzed a sample of 65 empirical studies conducted from 1971 to 1985 on evaluation utilization. These studies utilized a variety of methodologies including retrospective, longitudinal, and simulation research designs. Cousins and Leithwood (1986) defined evaluation utilization or use in three different ways: as support for discrete decisions such as decisions about program funding, the operation of a program, or program management; as information used to educate decision makers; and as psychological processing of evaluation results. Cousins and Leithwood (1986) reviewed the methodological characteristics, dependent variables, and independent variables of empirical studies in education, mental health, and social services to identify factors that influenced the

use of evaluation results. Similar to Leviton and Hughes (1981) they neglected to articulate how these factors were mediated by how the *user* was defined in the studies, choosing instead to focus on the user as the person who adopted or made the decision to conduct evaluation.

The authors identified twelve factors that influenced one or more types of use. Six factors were highlighted as implementation of evaluation issues and consisted of evaluation quality, credibility, relevance, communication, the findings themselves, and the timeliness of evaluations for users. The other six factors were related to features of decision or policy setting such as information needs of users, decision characteristics, political climate, competing information, personal characteristics of users and user commitment and receptiveness to evaluation information. Based on these twelve factors, Cousins and Leithwood (1986) developed a framework suggesting relationships between the factors as they relate to implementation and setting influencing evaluation utilization, though they made no attempts to hypothesize which factors would have a greater impact on use of evaluation. Cousins and Leithwood (1986) drew six major conclusions from their meta-analysis and asserted “evaluation use seemed to be most evident when:

- Evaluations were appropriate in approach, methodological sophistication, and intensity;
- The decisions to be made were significant to users and of a sort considered appropriate for the application of formally collected data;
- Evaluation findings were consistent with the beliefs and expectations of the users;
- Users considered the data reported in the evaluation to be relevant to their problems;

- A minimum amount of information from other sources conflicted with the results of the evaluation” (p.360)

Based on these findings, one approach that Cousins and Leithwood (1986) advocated was the importance of engaging evaluation users, which they seemed to define as those who made the decision to conduct evaluation, throughout the evaluation process and using Cooley’s (1983) “client-oriented” evaluation method. This, Cousins and Leithwood (1986) argued, could improve credibility, relevance and increase commitment to the evaluation process by ensuring evaluation is responsive to users needs. In contrast to this, in a three case study that utilized Cousins and Leithwood’s (1986) factors that affect utilization, Levin (1987) found that contextual factors were more important in explaining use.

Building on this earlier work (Cousins & Leithwood, 1986), Shulha and Cousins (1997) conducted a review and synthesis of all literature (empirical and theoretical) on evaluation use, regardless of user, published since 1986. While this synthesis was weakened by their lack of differentiation between empirical and conceptual research, their lack of clear definitions of *use* and *user*, and not as extensive or methodologically sound as the earlier study by Cousins and Leithwood (1986), Shulha and Cousins (1997) attempted to clarify developments in theory, research, and practice. The main developments highlighted by Shulha and Cousins (1997) are based on conceptual research and included the recognition that organizational context was a critical factor that impacts use,⁵ that conceptions of use

⁵ The growing importance of context in utilization theory was influenced by the “Weiss-Patton” debate at the 1987 AEA meeting followed by a series in *Evaluation Practice* (Patton, 1988; Weiss. 1988a, 1988b). While Patton (1988) positioned the evaluator as responsible for advocating use, Weiss argued that evaluators can not always be held accountable as competing structural and political factors are at play in some decision making contexts that may have a greater impact on whether evaluation is used.

need to be expanded to include misutilization,⁶ process use and use at both the individual and organizational level, and diversification of the evaluator to a role of facilitator, planner, and educator. While this research is predominantly focused on a case study approach or evaluator's insights, it conceptually reinforces the importance of understanding the structural factors that may either facilitate or impede multiple stakeholders from using evaluation information.

Summary of Empirical Work on Evaluation Use Prior to 1996

The main finding from three extensive syntheses reviewing over 120 studies on evaluation use by Leviton and Hughes (1981), Cousins and Leithwood (1986) and Shulha and Cousins (1997) was the lack of empirical data in this area, as much of the research hinged on retrospective statements and individual case studies. Understanding these limitations, these reviews revealed a few generic conclusions. All of the authors emphasized the importance of understanding the context in which evaluation is implemented, in addition to understanding that who is involved in the evaluation process, how *use* (and *user*) are defined, and how this process is communicated to stakeholders can impact multiple dimensions of use: instrumental, conceptual, symbolic, and process and impact change at multiple levels: individual, group, and organizational.

⁶ Misutilization of evaluation can be defined as misuse of commissioning the evaluation, misuse of the evaluation process, and misuse of the findings.

Review of Empirical Work on Evaluation Use Since 1996

There are a number of recent studies that discuss evaluation use from the “insider perspective” and offer reflective insights into factors surrounding evaluation use based on specific cases. However, as Cousins et al. (2004) note, while these papers are “rich in detail and insight, it is not possible for the analyst to judge the credibility or trustworthiness of the account on the basis of conventional canons for social inquiry” (p.108). Consequently, this section of the review focuses on recent empirical work that clarifies methods, uses a sample, or describes other relevant sources of evidence. To assess whether the empirical literature has moved beyond earlier empirical work that was mainly case studies and retrospective accounts, the research will be divided by how the authors defined their methodology: meta-evaluation, case study and mixed methods research. As there is limited literature in this area, all literature will be included, regardless of whether it focuses on the nonprofit or for-profit sector, in addition to an older related research study on knowledge utilization. This will help to identify where there are gaps in the literature, particularly in regards to systematic methodological approaches and study populations. An overview of these studies is provided in Table 4.

Table 4

An Overview of Empirical Research on Use of Evaluation

Study	Sample	Context	Methods	User	Use Measure	Purpose/Research Question	Main Findings
Avery & VanTassel-Baska (2001)	Multiple staff at two gifted program evaluations	2 private K-12 schools	Retrospective case study approach; interviews	Program staff	Not defined	What factors facilitate the use and non-use of evaluation results by program staff?	Engaged program staff who supported evaluation found that lack of supportive leadership and other resources impeded use of results
Balthasar & Rieder (2000)	Evaluators, program directors and federal office of energy decision makers from 4 out of 40 evaluations	Evaluation of the Swiss Energy 2000 program	Meta-evaluation: 8 interviews plus document review in each of 4 case studies	Decision-makers	Direct and indirect use at three levels: operation, structure and strategy	-The forms of learning through evaluation -The conditions of learning through evaluation	Different kinds of learning in each case study: direct—where feedback information led to the recommendations of the evaluation being used; indirect—new appreciation of the situation is observed and arguments can influence political issues
Bedell, Ward, Archer & Kirk Stokes (1985)	213 employees from 30 different nonprofits	Mental health agencies trying to increase use of treatment innovations	Structured interviews; multivariate analysis	Program staff	Louck et al.'s (1975) level of use	Investigating relationships between eight theoretical factors of knowledge utilization and the implementation of treatment innovations	Four factors that predict knowledge utilization and the implementation of treatment innovations: value, idea, obligation and yield

Table 4 (continued)

Study	Sample	Context	Methods	User	Use Measure	Purpose/Research Question	Main Findings
Matlay (2000)	6000 organizations from yellow pages; 600 small business managers; 60 matched case studies	A report of a study of organizational learning (OL) in the small business section of the UK economy	Field study: multi-method staged study involving telephone survey, interviews, case studies	Managers	Not defined	How do small businesses use evaluation internally? What types of informal/formal organizational learning processes are in place?	Mostly informal learning takes place. OL occurs more with decisionmakers, such as managers/owners. Firms that employed intentional learning strategies were slower and more deliberate, but more effective at influencing the direction and delivery of a firms' programs
Robinson & Cousins (2004)	9 Canadian key informant national program practitioners	Canadian national training program reform initiative	Longitudinal case study; participant observation, interviews, focus groups	Multiple stakeholders internal to the organization	Not defined	To investigate the organizational consequences of internal participatory evaluation in a Canadian national training organization	Significant organizational consequences observed, especially capacity building. Organizational climate enabled participatory evaluation
Russ-Eft et al. (2002)	23 sales people, 10 implementation specialists, 9 consultants, 40 representatives of client companies	Evaluation of a sales program within a business services organization in the U.S.	Descriptive case study; satisfaction interviews; surveys	Sales program staff	Process use	Factors contributing to use and non-use of evaluation results	Program terminated on the basis of non-evaluative information; process use: developed shared understanding, engagement and self-determination.

Table 4 (continued)

Study	Sample	Context	Methods	User	Use measure	Purpose/Research question	Main findings
Scribner et al. (1999)	35 school employees (school admin, leadership team members and other teachers) from 3 rural middle schools	U.S. district-wide initiative	Collective case study; school-based observation; document and artifact collection	School employees	Not defined	How do school improvement programs foster the development of professional communities? What organizational factors facilitate or impede this process?	More stakeholders need to be involved in design, implementation and evaluation of school improvement programs. Factors influencing development include: leadership, past events, politics of allocating resources, school organization and teacher workload
Vakola (2000)	3 construction corporations	Organizations in construction industry in 3 European countries that underwent a business process re-engineering exercise	Descriptive case studies; data from semi-structured interviews	Leadership; high level decision-makers	Not defined	Is there a link between evaluation activities, organizational learning and innovation?	Evaluation resulted in positive benefits: facilitated the change process, led to better OL (not defined), led to process innovation, such as the introduction of improved technological capabilities in the companies.
Valovirta (2002)	27 agency and supervising department personnel	Government agencies in Finland	Meta-evaluation: semi-structured interviews	Management	Not defined	To compare how agency approaches to evaluation and how results were used for management development	Use takes place in social context: Types of arguments about evaluation results: used to legitimate existing practices, use for debate, written reports to promote dialogue

Research Using Meta-Evaluation

Balthasar and Rieder (2000), defining their study as a meta-evaluation in Sweden, conducted a document review of four government program evaluations and eight in-depth interviews with evaluators, program directors and other decision-makers. The purpose of this study was to analyze linkages between evaluation capacity building and use, defined as program directors and other decision-makers using evaluation results for direct and indirect learning. Based on their findings, Balthasar and Rieder distinguished between different types of evaluation use and learning at three different levels: operation, structure and strategy. At an operational or program level, program improvements are made; at a structural level, factors that will facilitate these program improvements or improve delivery of services are included; at a strategic level, the underlying assumption and goals of the program are challenges and reconceptualized. A major weakness of this study is related to the small number of interviews and the fact that the study only reviewed four of a population of forty programs that had been evaluated.

Valovirta (2002) conducted a meta-evaluation in Finland, used semi-structured interviews with 27 government agency staff to assess how policymakers and supervising department personnel used evaluation results. Valovirta (2002) did not provide a definition of use, but noted that evaluation use is most often used indirectly for argumentation purposes in social contexts. Examples of this documented by Valovirta (2002) include policymakers and supervising department personnel using data to legitimate a previous action, using evaluation to incite debate around a particular issue or program deficiency, and similarly, the use of written reports such as “white papers” to promote dialogue. Valovirta (2002)

concluded that these examples highlight how evaluation contributed to systemic capacity building and decision making processes. What was not discussed was what factors affected use by policymakers and supervising department personnel and how the evaluation results benefited the actual programs being evaluated, rather than simply being used by high level decision-makers.

Case Study Approaches

In a longitudinal case study based on participant observations, interviews, and focus groups with nine school training program practitioners in Canada, Robinson and Cousins (2004) investigated how schools used evaluation to improve programs following internal participatory evaluation processes. Robinson and Cousins (2004) concluded from their data that the participatory evaluation process resulted in a significant reconceptualization of the training program and how to build organizational capacity to improve program delivery. A major weakness of this study was that it did not discuss which factors impacted use and was based on a sample of only nine training program practitioners in one field within the nonprofit sector and therefore could not be generalized to other nonprofits.

Preskill, Zuckerman, and Matthews (2003) conducted a systematic case study of two evaluations that included sixteen interviews with advisory group members of the American Cancer Society (a nonprofit) engaged in the evaluation process. Preskill et al. (2003) wanted to ascertain what advisory group members learned from participating in the evaluation process and to identify factors that facilitated or impeded them learning from evaluation results. The authors noted several process uses connected to being involved in the

evaluation process, including learning more about the evaluation process and the program itself. Preskill et al. (2003) identified several factors that impacted these process uses: management support, communication networks, and individual and organizational level characteristics, such as an organizational culture that supported learning. Again, while this case study was conducted systematically, it focused on how advisory groups used evaluation and did not actually reference any changes occurring in the programs or organizations as a result of evaluation.

In the for profit sector, Vakola (2000) used descriptive case studies and multiple semi-structured interviews with three construction organizations engaged in a restructuring process to assess the linkages between evaluation activities, the use of results (mainly by leadership) for organizational learning, and innovation. Vakola (2000) concluded that use of evaluation facilitated the innovation process outcome and helped implement organizational change by identifying factors that would enable improved capacity within the organization. Similar to the findings by Preskill et al. (2003), Vakola (2000) also noted the important role of supportive leadership and organizational characteristics in the evaluation and organizational change process. As this is simply a descriptive case study, any findings must again be treated cautiously as there is insufficient methodological information to evaluate this study fairly.

Also in the corporate sector, Russ-Eft, Atwood, and Eggherman (2002) explored use and non use of evaluation results in the context of a sales program within a business services organization. Russ-Eft et al. (2002) did not provide a clear definition of use and used a descriptive case study approach and interviews and surveys with approximately 82

stakeholders connected to a sales program. These stakeholders included 23 sales people, 10 implementation specialists, 9 consultants, and 40 representatives of client companies. While their predominant finding was that external forces, such as politics, were more important factors than evaluation results in deciding whether the program continued or not, Russ-Eft et al. (2002) did note that evaluation was useful to program staff in several process ways: to develop shared understanding, improve engagement, and promote ownership of the evaluation process.

Using a retrospective case study approach based on interview data with multiple staff at two gifted program evaluations in K-12 schools,⁷ Avery and VanTassel-Baska (2001) found that structural factors, such as lack of leadership support, time, staff, and other resources, impeded the use of data by program staff. While program staff acknowledged that the evaluation was useful, raised consciousness about the programs and practices and could be used as a catalyst for program and organizational change, there were barriers, such as lack of available resources and lack of leadership support, which stymied programmatic change efforts. While there are weaknesses in this study in terms of the lack of a systematic approach and the use of retrospective interview data, Avery and VanTassel-Baska (2001) did highlight that even when program staff are engaged and support evaluation, there are other factors, such as lack of supportive leadership and other resources, which may impede use of evaluation results by the organization.

Using three rural middle schools, Scribner, Cockrell, Cockrell, and Valentine (1999) used a collective case study approach and looked at the organizational impact of using

⁷ While this was not explicitly stated, the majority of schools are designated nonprofits

evaluation results from school improvement programs. Based on document collection and input from 35 teachers and school administrators, Scribner et al. (1999) concluded that the most important factor that impacted programs using information for organizational development was the leadership style of the principal. Other factors linked to the leadership style of the principal included the organizational climate, the school and principal's past experiences with organizational change, and whether there was a competitive model in place to allocate resources. Based on this information, Scribner et al. (1999) concluded it was unclear that the evaluation program information led to change due to the importance of the change agent, in the form of the principle, and existing organizational culture. While this study was based on only three case studies, it systematically used a comparative case study approach and reiterates the importance of organizational characteristics, such as supportive leadership and whether a climate exists that supports using evaluation for learning.

Mixed Methods Research

In the private sector, Matlay (2000) utilized mixed-methods, including 60 case studies, 600 interviews with small business managers, and a large-scale telephone survey using a national sample of 6,000 small businesses, to examine how organizations used evaluation internally, what types of organizational learning processes were in place, and how formalized these processes were. While this extensive study was based on businesses in the for-profit sector, it used a sample of small businesses, who share similar resource and time constraints as nonprofits. Matlay (2000) found that due to time constraints and the demands of leadership, the majority of learning was not based on evaluation results, was relatively

informal, and used in the short-term. Matlay (2000) further noted that this informal learning that did not use evaluation results did not build organizational capacity or contribute to improving organizations and sustaining competitiveness. In contrast, while companies that used more systematic data collection emphasized that it took more time and resources, there was a consensus that use of evaluation results by organization staff contributed to organizational improvements and growth.

Related Research: Knowledge Utilization

In an empirical evaluation of knowledge utilization, Bedell, Ward, Archer, and Kirk (1985) used a sample of mental health agencies (which are predominantly nonprofits) and conducted a retrospective investigation of the relationships between eight theoretical factors of knowledge utilization (Davis, 1971) and the implementation of treatment innovations. Bedell et al. (1985) developed a questionnaire based on Davis' (1971) evaluation utilization model that describes eight factors that are essential in determining organizational readiness for change: ability, values, idea, circumstances, timing, obligation, resistance, and yield (A-VICTORY) and divided use into three levels by adapting Loucks et al.'s (1975) level of use measure.

Based on structured interviews with 213 employees from 30 different organizations, Bedell et al (1985) conducted multivariate analysis of variance (MANOVA) and found significant positive relationships between four of the A-VICTORY factors: value, idea, obligation, and yield and increased use of treatment innovations. Based on Davis' (1971) descriptors of these factors, Bedell et al. (1985) concluded that congruence between the

values of an organization and those inherent in an innovation (value), ensuring stakeholders have sufficient information about an innovation (idea), the impetus the organization feels for action or change (obligation) and the organization's perception of the perceived benefits of adopting the innovation (yield) are critical factors that impact whether or not an organization will use an innovation.

While Bedell et al.'s (1985) research was based on innovation or knowledge utilization; it does ratify some of the research around evaluation use, particularly in regards to engaging stakeholders in the evaluation process, especially if evaluation is posited as a social innovation. If we translate this to evaluation in the nonprofit sector then it reinforces Klein and Sorra's (1996) emphasis that it is essential that there is congruence between what is being implemented (evaluation) and the organizational culture (value), that stakeholders understand the innovation process (idea), that the importance and need for this innovation (evaluation) is communicated throughout the organization and supported by leadership (obligation) and that all stakeholders perceive this innovation (evaluation) as useful to their organization (yield). This reemphasizes the importance of the organization having a climate conducive to using evaluation, of stakeholder engagement throughout the evaluation process, and of an organization's leadership being supportive of the evaluation process.

Summary of Recent Research around Utilization of Evaluation and/or Knowledge

The recent research on evaluation use continued to be dominated by case study approaches, with limited mixed-methods research, and no clearly defined large scale survey assessments that would allow for some form of comparability across organizations or

assessment of the level of impact of factors that impact use. This research makes little attempt to define use clearly and to emphasize the importance of identifying who is using the results, focusing mainly on policymakers or leader's use of evaluation rather than how it can be used to improve programs and organizations. In addition, the only research conducted in nonprofits is based on case studies in schools, which represent only one field of the nonprofit sector and are a special case as the majority of schools receive automatic nonprofit designation.

Despite these weaknesses, these case studies did document specific uses of evaluation that included a direct impact on decision-making, organizational improvements, increased understanding of both the evaluation process and the program or organization and contributing to overall organizational capacity building. In accordance with research prior to 1996, the research findings continued to emphasize the importance of involving multiple program and organizational staff in the evaluation process, as involving more stakeholders impacted both process use and direct use. In addition, the research since 1996 has brought organizational characteristics, such as an organization having an existing climate for learning, organizational values congruent with using evaluation, resources to support evaluation and supportive leadership to the forefront. These organizational characteristics are highlighted as significantly impacting whether and how evaluation results are used. This research supported an earlier study conducted by Bedell et al. (1985), which emphasized the importance of organizational culture and capacity, stakeholder engagement and supportive leadership to utilization of information.

Use of Evaluation: A Critique of the Research and Implications for Nonprofits

Despite the almost universal declaration of the importance of evaluation and theoretical work surrounding factors that impact multiple uses of evaluation, there is clearly limited empirical work on evaluation use or a clear definition of the construct of “evaluation use” and/or use by local program staff rather than high leveled decision-makers, such as funders. As Weiss (1988a, 1998) noted, there are barriers to evaluation use at the individual, group, organizational, and systems level and more research is needed to understand evaluation use as a multidimensional construct that incorporates both the process and product or findings of evaluation, at multiple levels.

Cousins et al. (2004), Leviton (2003), Torres and Preskill (2001), and Weiss (1988a, 1998) all emphasized the need for more systematic empirical work around evaluation use that moves beyond individual case study approaches so as to allow for some comparability of organizations. In her work on the advances, challenges, and applications of evaluation use, Leviton (2003) critiques the paucity of research surrounding evaluation use in addition to the quality, which Leviton (2003) notes “suffers from a flawed standard of evidence” (p.526). In contrast to the way in which evaluation is conducted, Leviton critiques the frequent use of self report with no validation of measurement or triangulation of information. Based on an analysis of 36 studies related to evaluation utilization and evaluation capacity building, Cousins, Goh, Clark, and Lee (2004) noted that research around evaluation use continues to be underdeveloped due to individual case study approaches and small sample sizes that reduce the power of any analysis. Of the 36 studies identified, Cousins et al. (2004) noted that fifteen were based on reflective accounts, usually by the evaluator, supported by

one or more specific examples, five were qualitative case studies, five were comparative case studies and six were based on longitudinal data collection or multi-method field studies that had limited methodological description. Cousins et al. (2004) concluded that the heavy reliance on qualitative methods and limited variability in modes of inquiry suggested that this literature is still at an early stage of maturity and consequently, findings should be treated cautiously. Cousins et al. (2004) emphasized the importance of quantitative studies that utilize latent structures, path analysis and hierarchical explanations to move the field forward.

Literature that discusses both the implementation and use of evaluation in the nonprofit sector is limited and reveals the challenges associated with this process. Poole et al. (2000, 2001) describe the lack of consensus by funders as to what actually constitutes evaluation, which in turn leads Carman (2005) to critique the limited agreement over the prevalence of evaluation practice across the sector. While some researchers point to improved evaluation practice (Billitteri, 1998; Wholey & Hatry, 2001), others suggest there are limited resources for evaluation purposes (Alie & Seita, 1997). The diversity and uniqueness inherent in the nonprofit sector in terms of mission and approaches toward interventions complicates the evaluation process further. Questions that have yet to be researched using samples of nonprofits include how evaluation is being approached, how (and whether) evaluation is being used by nonprofits and how evaluation can be used by nonprofits to improve programs and practices (Bozzo, 2000, 2002; Fine et al., 1998).

While foundations and other funders have developed resources to guide nonprofits through the evaluation process, the importance of capacity building to facilitate the

evaluation process is largely overlooked (Bies, 2001; Bozzo, 2000, 2002; Carman, 2005; Chinman et al., 2005; Kirsch, Krupa, Horgan, Kelly, & Carr, 2005; Poole et al., 2001).

Almost no research has looked at the extent to which organizational factors such as size, leadership, foundation support and a culture of learning impact the implementation and use of evaluation by the nonprofit (and not just for accountability). This information is critical if we are to develop the capacity of the nonprofit sector to use evaluation to improve service delivery, and ultimately, to benefit our communities. Given that this study's focus is on use of evaluation by nonprofits, which can be characterized as using evaluation for organizational learning, rather than for accountability to external stakeholders, a review of the research surrounding the construct of organizational learning is necessary.

Organizational Learning

To identify literature that discussed the role of organizational learning in relation to nonprofits or organizations using evaluation, I used three approaches. Initially, I conducted a computerized bibliographic search of databases including Business Source Premier, LexisNexis Academic, Web of Science (Social Sciences Citation Index) and PsycINFO using the keyword organizational learning, refined by combinations of the keywords organizational learning, evaluation, nonprofit, learning capacity, organizational change/development, change management, transformation, and systems. I then searched the bibliographies and reference lists of these articles for additional references. Finally, I scoured the nonprofit and evaluation journals using the same list of keywords in combination. While it would have been useful to find studies that dealt with samples of nonprofits, in the few studies that did

not use organizations from the for-profit sector, it was not made apparent whether organizations that could potentially be nonprofits (such as community agencies) were registered nonprofits.

Empirical Research around Organizational Learning

Research on evaluation use by organizations, rather than by external stakeholders, highlights various factors that may be important, such as stakeholder involvement, leadership and an overall environment that promotes learning within an organization. These factors are consistent with the theory around the construct of organizational learning, which suggests that organizational learning occurs when an organization has the skills necessary to adapt its behavior by transferring and reflecting new knowledge (Garvin, 1993). Theoretical work by Senge (1990a, 1990b), Garvin (1993), Birleson (1998), Torres, Preskill, and Piontek (1996) and Watkins and Marsick (1996) suggests that for the process of organizational learning to happen, an organization must have certain characteristics in place, such as supportive leadership fostering an environment conducive to learning at all levels of the organization, which is informed by the perspectives of multiple stakeholders. Collectively these factors can be described as organizational learning capability, which predicts the latent factor of organizational learning.

To look more closely at the linkages between organizational learning capability and the positive impact on programs and practices, several researchers have taken on the challenge of operationalizing organizational learning capability and establishing relationships between the underlying organizational learning characteristics and proxy measures of organizational learning, such as organizational performance. This is a relatively new area of

research that has been developed within the for-profit sector and consequently, the organizational setting will be made explicit throughout this section. The majority of these studies are based in the for-profit sector and have relied on survey data and financial performance measures, though a few studies have started to move beyond financial performance and look at the linkages between organizational learning, innovation, evaluation, and multi-leveled change, both direct and indirect. This section is divided into research that uses specific instruments that have been developed in relation to organizational learning capability and provides a description of the development of each instrument, along with the proxy measures of organizational learning, such as organizational performance.

The Dimensions of the Learning Organization Questionnaire (DLOQ)

Watkins & Marsick (1993, 1996) developed the Dimensions of the Learning Organization Questionnaire (DLOQ), which is based on seven complementary action imperatives they identified as important in enabling the latent construct of organizational learning: continuous learning, dialogue and inquiry, team learning, empowerment, embedded systems, system connection, and strategic leadership (Figure 7). Watkins & Marsick (1993) measured these seven dimensions using 43 items and assessments of the psychometric properties suggest that these dimensions have acceptable reliability estimates (coefficient alpha ranges from .75 to .85) and the seven factor structure fit the empirical data well enough to be appropriately applied in the context of organization research (Yang, Watkins, & Marsick, 1998; 2004).

Action Imperative	Definition
Create continuous learning opportunities	Learning is designed into work so that people can learn on the job; opportunities are provided for ongoing education and growth.
Promote inquiry and dialogue	People gain productive reasoning skills to express their views and the capacity to listen and inquire into the views of others, the culture is changed to support questioning, feedback and experimentation
Encourage collaboration and team learning	Work is designed to use groups to access different modes of thinking, groups are expected to learn together and work together; collaboration is valued by the culture and rewarded
Establish systems to capture and share learning	Both high-low technology systems to share learning are created and integrated with work; access is provided; and systems are maintained
Empower people towards a collective vision	People are involved in setting, owning and implementing a joint vision; responsibility is distributed close to decision making so that people are motivated to learn what they are accountable for
Connect the organization to its environment	People are helped to see the impact of their work on the entire enterprise; people scan the environment and use information to adjust work practices; organization is linked to community
Leaders model and support learning	Leaders model, champion and support learning; leadership uses learning strategically for business result.

Figure 7: Seven actions that promote organizational learning (Watkins & Marsick, 1993, 1996).

Ellinger, Yang, & Ellinger (2000) conducted an exploratory mail survey study to look at the relationship between the dimensions purported to predict learning organization and for-profit firm performance using Watkins & Marsick's DLOQ (1993) along with objective measures of firm financial performance, such as return on equity (ROE) and return on assets (ROA). Using a random sample of 400 mid-level managers at U.S. manufacturing firms, Ellinger et al. (2000) used canonical correlation to assess the omnibus impact of the dimensions of the learning organization on a set of perceptual (managers perceptions) and

objective (financial) measures of firm performance. Ellinger et al. (2000) did not discuss the structure of the variables but focused on the overall effects of the canonical correlation analyses.

Ellinger et al. (2000) found statistically significant canonical correlation effect sizes (.246 to .312), which indicated that more than twenty-five percent of the respondents' perceptions of organizational performance could be accounted for by the seven dimensions of the learning organization. The canonical correlation between the seven dimensions of the learning organization and the four secondary measures of financial performance (.104 to .108) further supported this positive association, as more than ten percent of the variance in the four financial indicators was explained by the DLOQ. While this exploratory research strengthened the credibility of the organizational learning concept, weaknesses of the study included a bias toward larger, publicly owned firms, the use of only four measures of financial performance, and the focus on mid-level managers to the exclusion of upper-level managers or front-line employees.

This attempt to interlink learning organization characteristics and firm performance was replicated in a Chinese for-profit context by Zhang, Zhang, and Yang (2004), who used the DLOQ (Watkins & Marsick, 1993) to look at the association of learning organizations with firm performance and the differences between service versus manufacturing industries. Zhang et al. (2004) conducted a survey of 477 middle-level managers in six different firms. Two perceptual measures of organizational performance (finance and knowledge) were used as dependent variables and the DLOQ were used as predictors. Using canonical correlation analysis, Zhang et al. (2004) found a statistically significant effect size of 0.11 to 0.12, which

suggested that more than 10 percent of the variability in performance could be accounted for by the seven dimensions of the learning organization. This was lower than the effect size for perceptions of performance found by Ellinger et al. (2000), which the authors attributed to the more complicated nature of Chinese enterprises during the economic transition process. Independent sample t-tests showed that service companies exhibited better learning practices than manufacturing companies, but the authors did not conduct further analysis to assess whether this impacted the relationship between learning practices and performance. Consequently, while this research further strengthened the underlying dimensions of organizational learning, the limitations included using a convenience sampling method that consisted of only mid-level managers, who potentially have differing perspectives than others in the organization, and the fact that the study relied on perceptual performance measures rather than actual financial data. While this is based on the for-profit sector, the stronger relationship between the dimensions of organizational learning and performance in the service sector could potentially be applied to the predominantly service-oriented nonprofit sector and highlight the relevance of organizational learning characteristics to service-oriented organizations.

Taking a slightly different approach, Ismail (2005) conducted a mail survey study to explore the effects of creative climate and organizational learning dimensions on innovation with 259 employees at 18 different for-profit organizations in Kuala Lumpur; thirteen were locally owned and five were multinational corporations (MNCs). Ismail (2005) used a self-report questionnaire that contained three constructs: learning culture dimensions measured by Watkins and Marsick's 43-item DLOQ (1993); creative climate measured by the creative

climate questionnaire (CCQ) (Ekvall, Arvonen, & Waldenström-Lindström, 1983); and a 32-item innovation measure that contained two constructs: technological transfer/diffusion of innovation and organizational innovation, based on guidelines by OECD (1997) and MASTIC (1996). Ismail (2005) conducted reliability estimates for each measure and found the reliability estimates for the CCQ was .94, DLOQ was .97, and innovation was .98. Using multiple regression analysis, findings suggested that independently, the DLOQ accounted for about 57 percent of the variance around innovation and made a more significant contribution toward explaining innovation than the CCQ, which accounted for 35 percent of the variance. Ismail (2005) reported that factor analysis using Varimax rotation with Kaiser Normalization revealed the two constructs were not highly correlated with each other. Collectively, regression analyses revealed that the joint contribution of both the DLOQ and the CCQ explained 60.2 percent of the variance around innovation in the 13 local organizations and 68 percent of the variance around innovation in the MNCs. Parsing out the factors underlying each construct, Ismail (2005) concluded that creating a climate for learning had a greater impact on innovation than any of the creative climate factors. The main limitation of this research is linked to the sampling design. While Ismail (2005) used a random sample of private organizations, the employees were individually selected by the organization's representatives, potentially biasing the positive responses toward the mail survey. Despite this potential bias, the correlations between being a learning organization and being innovative do provide further support for the advantages of an organization having organizational learning characteristics.

In one of the few studies on nonprofits, Glosson's (2002) dissertation used Watkins & Marsick's DLOQ (1993) as a dependent variable as part of a qualitative case study to help evaluate, describe, and interpret the impact of the Human Dynamic program, an organizational learning program, across 26 sites of a healthcare organization. The Human Dynamics Program is an organizational development strategy that is promoted as a foundational approach to support the creation of a learning organization (Senge et al., 1994; Seagal & Horne, 1997). While the process is unclear, Glosson (2002) used four data collection methods: a survey of 105 past participants, six individual semi-structured interviews, two focus groups, and personal observations and stated that she used a rating code to validate the DLOQ and self-reported changes for these 105 respondents. Glosson (2002) concluded that two stages of field-testing indicated that the DLOQ was a useful tool to measure change in learning at the individual, team, and organizational level and found that the program resulted in five major improvement categories: self-awareness, understanding others, communications, team behaviors, and meeting facilitation. Glosson (2002) cited that the positive impact of individual learning behaviors can be leveraged to increase learning throughout the healthcare organization. What was not discussed was whether this program resulted in any tangible benefits for the healthcare organizations, beyond the individual level behaviors reported. In addition, this study is seriously hindered by the lack of methodological description and analysis of the constructs, in particular the problematic measure of retrospective self-reported changes.

In the Australian nonprofit sector, McCaffrey (2004) also used a mail-in adaptation of the DLOQ (Watkins & Marsick, 1993) to examine the impact of a management-training

program on the association between learning organization practices and organizational performance. One of the proposed outcomes of this management-training program was assisting organizations to learn and both learning organization practices and organizational performance were measured using the adapted DLOQ. McCaffrey's population of interest was 237 managerial employees from four public nonprofit organizations; these employees may or may not have participated in this program. Based on t-tests and Pearson correlations, McCaffrey (2004) concluded that there was a positive association between learning organization practices and the organizational performance measure in the public sector. In addition, McCaffrey (2004) found that managerial employees who had participated in the training rated their organizations significantly lower on effective learning organization practices. In terms of methodological weaknesses, more sophisticated data analysis that moved beyond t-tests to regress learning organization practices on performance, or controlled for other factors such as length of time in management, would have helped bolster the adequacy of McCaffrey's conclusion. Despite this, McCaffrey's work supported the findings of Ellinger et al. (2000) and Zhang et al. (2004) and reemphasized the presence of activities underlying organizational learning and the impact of organizational learning on performance.

Collectively, the multiple settings, both in the for-profit and nonprofit sector, for the research using the DLOQ demonstrate there are measurable dimensions underlying the construct of organizational learning. Ellinger et al. (2000), Zhang et al. (2004), and McCaffrey (2004) all emphasized the importance of the dimensions of organizational learning when they found positive associations between these dimensions and performance. Ismail (2005) further noted the importance of an organizational learning culture to

innovation. While these studies are limited by sample size, sampling design, and limited data analysis, the overall dimensions of the construct of organizational learning seem solid and there is increasing support for the importance of organizational learning in terms of its impact on performance. The major limitation of an instrument such as the DLOQ is its length (43 items), which makes it difficult to include as a scale of organizational learning in any mixed use survey that seeks to measure beyond simply the construct of organizational learning.

The Assessing Learning Culture scale

Botcheva, White, & Huffman (2002) developed the Assessing Learning Culture scale to examine learning culture and outcomes measurement practice as indicators of community nonprofits' readiness for conducting research-based evaluation. This 10-item scale is based on the important aspects of learning culture outlined by Preskill & Torres (1999), such as a focus on program and organizational processes as well as outcomes, collaboration, cooperation, participation, and shared individual, team and organizational learning. Using a convenience sample of 25 nonprofit community organizations, Botcheva et al. (2002) mailed out a survey that looked at the interrelationship between evaluation activities such as data collection practices, learning culture, and agency funding. Using correlational analyses, Botcheva et al. (2002) found a significant and positive correlation between learning culture and data collection practices, suggesting that a set of underlying beliefs and attitudes enable organizations to integrate data findings into their nonprofit and learn from their own experiences. Nonprofits that integrated data successfully were more able to use data to identify strengths and weaknesses within their programs, improve upon programs, adapt

practices and secure higher levels of funding. While the small sample size results in low statistical power, this does draw attention to the importance of learning culture as an important factor in conducting and using evaluation within nonprofits.

Summary of the Research on Organizational Learning

In order to establish relationships between organizational learning characteristics and organizational performance, researchers have developed instruments and conducted studies to look at the linkages between the dimensions of organizational learning, innovation, evaluation, multi-leveled change and performance. These can loosely be divided into research looking at linkages between predicted organizational learning dimensions and organizational performance and research looking at organizational learning dimensions and current evaluation practices.

Ellinger, Yang, and Ellinger (2000), Zhang, Zhang, and Yang (2004) and Ismail (2005) empirically tested the dimensions of organizational learning using Watkins and Marsick's Dimensions of the Learning Organization Questionnaire (DLOQ) (1993) in the for-profit sector and found positive associations between the dimensions of organizational learning and performance and innovation. Glosson (2002) and McCaffrey (2004) adapted Watkins and Marsick's Dimensions of the Learning Organization Questionnaire (DLOQ) (1993) for use in the nonprofit sector and McCaffrey (2004) concluded that there was a positive association between learning organization practices and organizational performance. One major limitation for the DLOQ however, is that it has 43 items, making it difficult to include as simply a scale of organizational learning in a survey including other variables.

Research looking at the linkages between organizational learning dimensions and current evaluation practices in the nonprofit sector was conducted by Botcheva, White and Huffman (2002). Botcheva, White and Huffman (2002) developed a 10-item instrument based on Preskill and Torres' (1999) work, and found a significant and positive correlation between the presence of a learning culture and how nonprofits conducted and used evaluation to benefit their organizations, improving upon programs and practices and securing higher levels of funding.

While there are some limitations to the research surrounding the organizational characteristics that predict organizational learning and the impact on performance and evaluation practice, there is clearly some support for the construct. This suggests the need for the inclusion of a measure that looks at the dimensions underlying organizational learning in any research that seeks to look at an organization's ability to use information internally, such as evaluation findings. In addition, while the research surrounding the organizational characteristics that predict organizational learning is an emerging literature, it would benefit from including more than bivariate correlations such as the relationship between organizational learning characteristics and more complex measures of performance or use. What remains unaccounted for in the literature is whether the organizational characteristics that predict organizational learning will continue to have an impact on factors, such as performance or use of evaluation, when other potentially significant and correlated factors, such as stakeholder engagement and supportive leadership are included in the model.

Stakeholder Engagement

To identify literature that discussed the role of stakeholder engagement in relation to nonprofits or evaluation, I used three approaches. Initially, I conducted a computerized bibliographic search of databases including Business Source Premier, LexisNexis Academic, Web of Science (Social Sciences Citation Index) and PsycINFO using combinations of the keywords stakeholder, evaluation, nonprofit, participatory, collaborative, and accountability. I then searched the bibliographies and reference lists of these articles for additional references. Finally, I scoured the nonprofit and evaluation journals using the same list of keywords in combination.

Stakeholder Theory

Stakeholder theory is based on the concept that organizations consist of a variety of stakeholders and should be managed with these stakeholders in mind (Greenwood, 2001, p.32). The credited founder of stakeholder theory, R. Edward Freeman (1984) suggested that systematic attention to all stakeholder interests is critical to any organization's success. The two main approaches to stakeholder theory are instrumental and normative. An instrumental approach argues that responsiveness to stakeholders leads to improved performance while a normative approach is based on ethical principles (Barrett, 2001; Cadbury, 1998, 1999). While much of the research is based on the for-profit sector, Barrett (2001) built on Carver and Carver's (1996) notion of moral ownership and argued that nonprofits need to be accountable to their stakeholders. Barrett used a case study approach to test the applicability of stakeholder theory to the nonprofit sector and concluded that despite nonprofit

stakeholders being different from for-profit stakeholders, it is possible to examine nonprofits using stakeholder theory, to describe stakeholders and their interest in the nonprofit, and to assess how the nonprofit manages its relationships with stakeholders and thus assess how accountable the nonprofit is to its stakeholders.

Based on Barrett's (2001) research that stakeholder theory is relevant to the nonprofit sector, a simple hub and spoke model of nonprofit stakeholders could be described as in figure 8. If we used this model for stakeholders engaged in the evaluation process, what would need to be added are others, such as external evaluators. In addition, internal staff could be divided into executive management, internal evaluators, and program staff to more clearly identify who is engaged in each part of the evaluation process. This model is used only to define a set of stakeholders and does not overlook the complex matrix of relationships that the various stakeholders have to each other, as well as to the nonprofit.

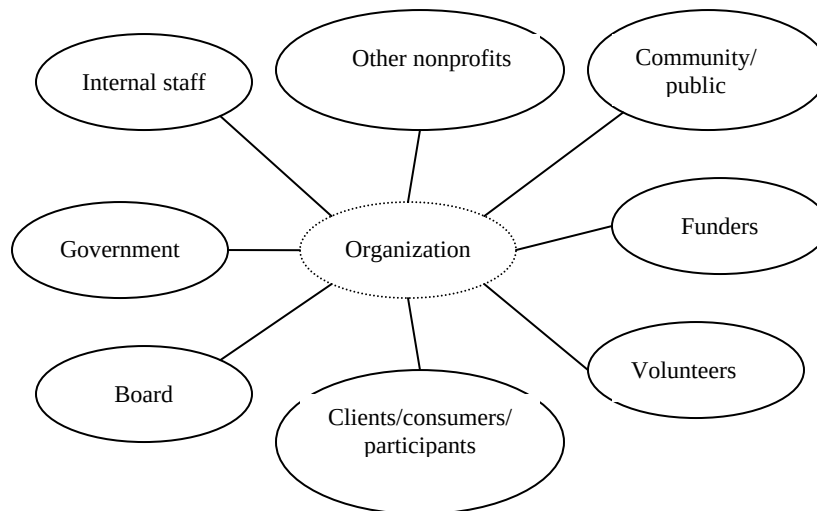


Figure 8: A hub and spoke model of the nonprofit stakeholders (Adapted from Freeman, 1984).

Empirical Research

Research by Brandon (1998), Fine, Thayer, & Coghlan (1998, 2000), Greene (1988a, 1988b), Turnbull (1999) and Taut and Alkin (2003) supported the linkages between stakeholder participation and improved evaluation processes. This is further validated by two extensive syntheses on evaluation use by Cousins and Leithwood (1986) and Shulha and Cousins (1997). These syntheses emphasized that stakeholder participation and the inclusion of diverse perspectives can stimulate knowledge production and promote a more positive environment for the implementation of evaluation and the use of evaluation findings by the organization (Shulha & Cousins, 1997).

Brandon (1998) conducted research in nonprofit educational settings to assess whether stakeholder participation could help enhance the likelihood of producing valid evaluation findings that would have a greater likelihood of being used by schools. Synthesizing four mixed-method research studies in educational settings that utilized procedures to improve evaluator-stakeholder interaction, Brandon (1998) analyzed observational and questionnaire data and concluded that increased stakeholder participation by school staff significantly improved evaluation validity. This was attributed to several factors, including: adapting methods based on stakeholder input, revising measures after conducting focus group sessions, and modifying evaluation recommendations before submitting a final report (Brandon, 1998). While Brandon's (1998) research is limited by the lack of methodological description or specific results, the author does help to outline specific ways in which stakeholders can contribute to and enhance the evaluation process, potentially making evaluation results more beneficial to school programs.

In the nonprofit sector, Fine, Thayer, and Coghlan (1998, 2000) and Thayer and Fine (2000) explored the role of stakeholder participation in program evaluation to gain greater insights into whether different levels of stakeholder participation are associated with organizational characteristics, evaluation characteristics, evaluation outcomes and the respondent's *perceptions* of the likelihood of the organization using evaluation. Using a three-phased mixed-method research design, Fine et al. (1998, 2000) collected a national sample of 178 mail surveys (a response rate of 59%) and conducted forty telephone interviews and four in-depth profiles to examine in detail the characteristics and outcomes of nonprofit responses. Weaknesses in the methodology and robustness of this study were that it relied on respondent's perceptions of the likelihood of using evaluation as a dependent variable, did not define use by whom, did not specify who was to complete the mail-in survey, relied predominantly on descriptive data and did limited testing for significance, used single item independent and dependent variables and grouped stakeholder participation into three equally numbered (rather than conceptually categorized) groups, which were defined as levels (low, medium, and high).

Fine et al. (1998, 2000) found that evaluations that included more stakeholders, such as program staff and not just executive directors and external stakeholders, were more likely (than in the low stakeholder participation group) to be used to improve outcomes or impact (92% compared to 64%), to promote the program to potential participants/clients (73% compared to 47%), to respond to questions/criticisms about the program (65% compared to 47%) and to decide resource allocation within the organization (58% compared with 22%). In contrast, low stakeholder participation evaluations were more likely to be primarily

designed for strategic planning (14% compared to 4%). Fine et al. (1998, 2000) further found that most respondents (70%) reported that stakeholder involvement played a key role in increasing the likelihood of programmatic changes being made based on evaluation findings and that respondents with high stakeholder participation found the evaluation results more credible and more satisfying than respondents with low-participation evaluations (65% compared to 45%).

Fine et al. (1998, 2000) also looked at nonprofits' perspectives in regards to what increases stakeholders' satisfaction with, and perceived usefulness and credibility of evaluation. This was measured using individual item self-report measures based on the perceptions of the nonprofit respondent. While Fine et al. (1998, 2000) assumed this was the executive director, no attempt was made to capture this and report self-report by whom. Fine et al. (1998, 2000) identified several factors that contributed to perceptions of an evaluation's usefulness including documentation of programmatic success, a useful program planning tool, and a focused evaluation design. Similar to findings by Brandon (1998), the authors noted that increased stakeholder involvement by internal program staff and participants, was credited with improving evaluation design (by ensuring relevant questions are asked and appropriate measures selected), increasing program staff and participants' understanding of the evaluation which improved implementation of the evaluation process and increased multiple internal and external stakeholders' understanding and appreciation of the agency.

While this research by Fine et al. (1998, 2000) was limited by its convenient (rather than conceptual or best-practice) groupings of stakeholder participation and by perceptions rather than descriptive measures of evaluation usefulness, it did offer some support for the

importance of stakeholder participation in evaluation process. Participants who reported higher use of stakeholders throughout the process perceived that evaluations were more useful, credible and satisfying to their stakeholders. Consequently, it could be important to include the level of stakeholder involvement when looking at factors that facilitate using evaluation. One improvement to the measure used by Fine et al. (1998, 2000) would be to ensure that if stakeholder groupings are used, they are based on a clearly defined framework of what determines a low versus high level of stakeholder involvement by outlining the importance of participation by internal stakeholders, such as program staff. Alternatively, a continuous measure of stakeholder involvement could be included to more clearly assess whether increased use of specific stakeholders, such as program staff, in the different stages of the evaluation process is associated with increased implementation and use of evaluation by the nonprofit.

Greene (1988a; 1988b) used a descriptive case study approach in the nonprofit sector to elicit and describe a set of dynamic linkages between stakeholder participation (including program participants, program staff, volunteers, executive management staff and board members) and data utilization based on qualitative data from two youth employment program sites (Figure 9).

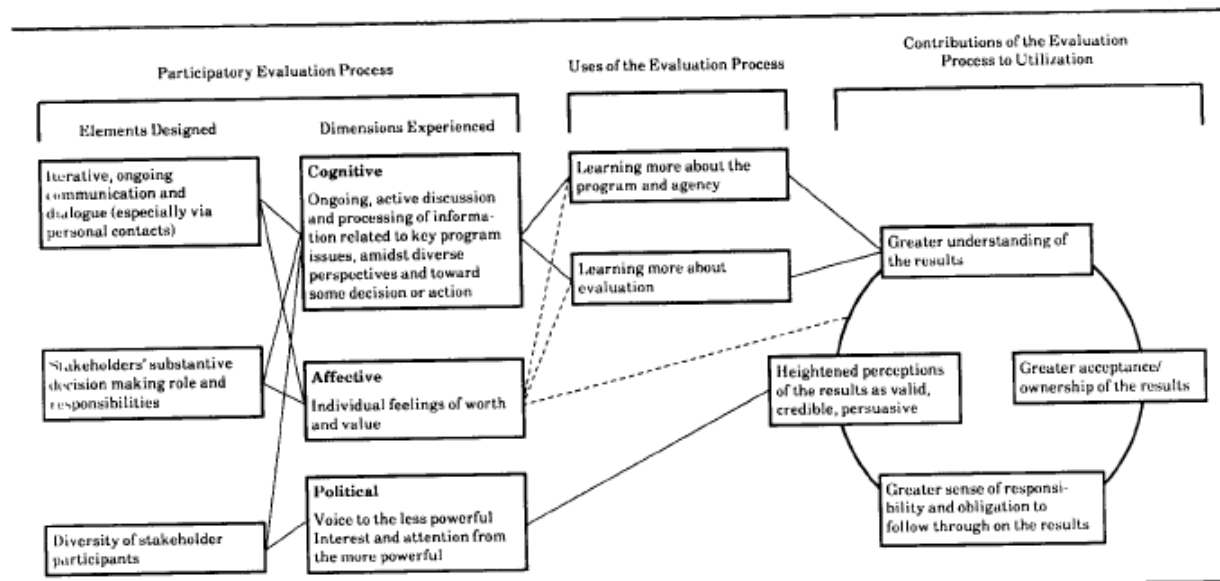


Figure 9: The dynamic links between stakeholder participation and utilization (Greene, 1988a, p.108).

As Figure 9 shows, key elements of this participatory process were linked to cognitive, affective and political dimensions of participation from the perspectives of these multiple stakeholders. These three dimensions were then linked to stakeholder attitudes and perceptions that predicted *potential* or *readiness* for utilization. Greene (1988a; 1988b) also documented actual use, both internal and external to the nonprofit, of evaluation findings as helping to increase participating stakeholder's program understanding, confirming and documenting intuitions about program effects, validating important directions for program development and planning, including data findings in funding proposals and enhancing program credibility within the community. These documented uses helped to develop a

typology of use that builds on Cousins and Leithwood's (1986) premise that processing is prerequisite to both conceptual and instrumental uses of results.

While Greene's (1988a; 1988b) methodological description was limited and little was offered by way of results, the major sources of data were: extensive documentation of the participatory procedures, methods, and instruments; stakeholder questionnaire responses; document reviews and post-evaluation interviews; and evaluator reflections (Greene, 1988a, p.100). One major weakness of the published study was that it did not offer any methodological insights into how the empirical grounding for this model (figure 9) was developed. This made it difficult to assess the robustness of Greene's (1988a) predicted model or use her model to replicate findings in any future research.

Critical to Greene's (1988a; 1988b) work is that *meaningful* stakeholder participation is defined as shared decision making. While stakeholders, such as program staff and program participants, can take part in the evaluation process by collecting data and consulting on design, this does not make them fully engaged in the evaluation process. This emphasizes the importance of clearly defining *level* of stakeholder engagement by looking at *who* is involved in *specific* aspects of the evaluation process.

While it was unclear how Greene (1988a) developed her model from the available data sources and she acknowledged that further research was needed to assess the participation-utilization linkages, her research pointed to several findings of interest in regards to the impact of wide stakeholder participation (figure 10). Greene (1988a; 1988b) argued that by substantively including a diverse group of stakeholders, from program participants to those in leadership roles, and providing opportunities to engage in dialogue

around evaluation information, multiple stakeholders understood the evaluation process more easily, felt more ownership of the results, and were thus more ready to accept evaluation findings, perceive them as credible and thus be more likely to use them to help their nonprofit. This is synonymous with the innovation paradigm, which suggests that involving multiple and diverse stakeholders in the innovation process at the earliest stages and sustaining this involvement, reduces the possibility of later rejection or reinvention of the implemented social innovation (Kanter, 1988; Klein & Sorra, 1996).

Also in the nonprofit sector, Turnbull (1999) used questionnaires with a sample of 315 elementary and secondary schoolteachers who had been engaged in an accreditation process to gain a better understanding of participatory evaluation by testing a set of causal relations in a proposed model of participatory evaluation. Using Chen's (1990) intervening mechanism design in conjunction with structural equation modeling, Turnbull (1999) proposed that a participative climate, level of participation in decision-making and influence in participative decision-making predict use of evaluation *by participants* through participation efficacy. Put more simply, Turnbull (1999) proposed that if participants perceive their work place as participatory, actively participate and believe this participation impacts the decision-making process, then it is likely that the evaluation process will also be perceived as effective and thus evaluation results will be used by those engaged in the process. Therefore, Turnbull (1999) hypothesized that increased participation in the evaluation process results in more positive perceptions of the effectiveness of evaluation, which in turn increases use of evaluation findings by those involved.

Using a time-lagged data collection process that utilized four questionnaires and a variety of Likert-type scales, Turnbull (1999) concluded that this proposed model was a plausible explanation of how participation can be expected to work to increase the use of evaluation information by participants, such as their sample of schoolteachers. Results showed that teachers who perceived their working environment to be participatory (participative climate) reported higher levels of participation and believed the evaluation process was effective as they felt they had a greater influence in decision-making. However, in contrast to the proposed model, results showed that while perception of being able to impact the decision-making process was significantly related to participation efficacy and consequently to self-reported higher levels of use by teachers, the relationships between evaluation use by teachers, a participative climate and higher levels of participation were statistically non-significant. Turnbull (1999) acknowledged that these insignificant relationships warrant further study, yet concluded that as the link between participation efficacy and use by teachers is significant, then her proposed model is substantiated.

Turnbull (1999) recognized that there are some questions as to the unexpected findings surrounding why only participation efficacy was significantly related to use by teachers and a participative climate and high levels of participation were not related to perceptions of evaluation effectiveness (and consequently use). However, Turnbull's (1999) study did offer some empirical support for the importance of stakeholders being involved in evaluation content decisions such as how evaluation is to be designed, analyzed, and reported and relates this to use of evaluation. This could support Greene's emphasis on meaningful stakeholder involvement as it emphasized that while level of participation and a participative

climate were considered important factors to participation, stakeholders still wanted to feel they had some influence on the decision-making parts of the evaluation process to feel that the evaluation had been successful. Consequently, any future studies looking at stakeholder engagement in the evaluation process should ensure they distinguish between the different aspects of the evaluation process and measure not only *who* is involved, but also whether they can impact any decision-making surrounding use of evaluation results. In addition, Turnbull's (1999) study reinserted the importance of a culture or context that has the capability to involve stakeholders in the decision-making process.

Taut and Alkin (2003) conducted a qualitative research study that, while not focused on stakeholder participation, noted its importance in relation to evaluation practice in general. In a series of interviews with eighteen staff members concerning program staff perceptions of barriers to evaluation implementation, Taut and Alkin (2003) categorized human factors as a major potential barrier to the implementation and use of evaluation by program staff. Taut and Alkin (2003) specifically noted program staff's lack of trust in evaluators and the evaluation process in general, which the authors further linked to inadequate relationship building. Given that stakeholder engagement plays a key role in relationship building in order to build trust in both the evaluator and the evaluation process, Taut and Alkin's (2003) work provided further empirical support for stakeholder engagement as a factor associated with effective implementation and use of evaluation.

Summary of the Theory and Research around Stakeholder Engagement

Stakeholder theory suggests that the active engagement of multiple stakeholders in decision-making processes can enhance an organization's success. While some take an instrumental approach and feel that including stakeholders is "a practical approach to broadening decision making and problem solving through systematic inquiry," others take a normative approach and feel that "reallocating power in the production of knowledge and promoting social change are the root issues (Cousins and Whitmore, 1998, p.5; Barrett, 2001; Cadbury, 1998, 1999).

While there is limited research surrounding how nonprofits use evaluation to benefit their organizations, several studies have started to build a knowledge base to empirically support the linkages between stakeholder participation and improved evaluation processes in the nonprofit sector. Brandon (1998) analyzed observational and questionnaire data and concluded that increased stakeholder participation by school program staff significantly improved evaluation validity, potentially resulting in increased use by schools. Fine, Thayer, & Coghlan (1998, 2000) and Thayer and Fine (2000) used a mixed methods approach and found that higher use of internal and external stakeholders throughout the evaluation process resulted in evaluations being perceived as more useful, credible and satisfying to the various stakeholders involved. Greene (1988a, 1988b) used a systematic case study approach to highlight that meaningful and diverse stakeholder participation, defined as shared decision-making, resulted in greater understanding and acceptance of evaluation results, increased perceptions of the evaluation as credible, and a stronger intent for stakeholders at multiple levels of the nonprofit to follow through and use evaluation results. Turnbull (1999) used

structural equation modeling to test a set of causal relationships in a model of participatory evaluation and found a significant link between effective participation and use of evaluation results by teachers and also highlighted the importance of a participative climate reminiscent of the characteristics underlying organizational learning.

Some of the limitations of this body of research also provide an opportunity for making contributions to this literature. Similar to the research surrounding organizational learning, none of this research has simultaneously looked at other factors that may facilitate improved evaluation processes, such as organizational learning and/or supportive leadership and whether this decreases the impact of stakeholder engagement. In addition, while the majority of this research discusses the importance of stakeholders being involved in decision-making and not just being a part of the evaluation process, little attempt has been made to assess how level of stakeholder engagement in decision making impacts the implementation and use of evaluation by nonprofit staff rather than high-level policymakers. Consequently, a study that looked more closely at whether increased use of stakeholders in the various stages of the evaluation process is associated with improved evaluation practices and a higher level of use by program staff would be a useful contribution to the field.

Role of Supportive Leadership

While the organizational learning literature stresses the role of management in promoting a learning culture and empirical studies on evaluation use by organizations have highlighted the importance of supportive leadership to use of evaluation results, there is no published literature that specifically theorizes and empirically examines the role of leadership

in relation to the implementation and use of evaluation. Despite conducting a computerized bibliographic search of databases including Business Source Premier, LexisNexis Academic, Web of Science (Social Sciences Citation Index) and PsycINFO using varied combinations of key words including leadership, leader, executive director, and evaluation, and searching the nonprofit and evaluation journals, no articles could be found that had empirically assessed the role of leadership in promoting the implementation and use of evaluation. Some of the most recent work related to the nonprofit sector has been presented at conferences, but not yet published. Consequently, this section will review a recent and available unpublished study and include references to previous studies that have simply emphasized that leadership emerged as a factor that impacted use of evaluation results.

As early as 1977, Patton, Grimes, Guthrie, Brennan, French, and Blyth conducted a qualitative study on evaluation use, primarily by those in decision-making positions, and interviewed leaders and evaluators from twenty federal health projects. Leadership emerged as a critical personal factor in the utilization process, as Patton et al. (1977) noted that if people in leadership positions supported and believed in the worth of the evaluation process, then results were more likely to be used in decision-making surrounding programs. Despite this, while others have made reference to interview data describing the importance of leadership in building capacity and promoting evaluation efforts (Carman, 2005; Levin, 1987; Leviton & Hughes, 1981; Plantz, Greenway, & Hendricks, 1997; Volkov and King, 2005), no quantitative work exists to support this or understand how the role of supportive leadership interacts with other factors to facilitate the implementation and use of evaluation by organizations. As Alaimo and Van Slyke (2006) highlighted, the nonprofit management

literature that relates to evaluation primarily focuses on barriers to evaluation, the program evaluation literature focuses on the engagement and skills of stakeholders, and both ignore the role of leadership in building internal capacity not only to conduct evaluation, but to make decisions about whether evaluation is used for accountability or for organizational learning.

Alaimo and Van Slyke (2006) proposed that program evaluation cannot be used to support organizational learning without the support and commitment of leadership. The authors explored this through a qualitative, multi-site, exploratory study interviewing a stratified random sample of 42 chief executive officers or executive directors of human service nonprofits in Atlanta, Georgia and Indianapolis, Indiana. Alaimo and Van Slyke (2006) transcribed and thematically coded semi-structured interviews to examine the role of nonprofit leaders and the extent to which they engaged in developing their organization's capacity for program evaluation.

In order to explore how leaders moved beyond accountability and promoted using evaluation for organizational learning, Alaimo and Van Slyke (2006) tentatively used Patton's (1997) definition of program evaluation that evaluation should be systematic, determine outcomes and use results to improve effectiveness to divide their nonprofit leaders into three levels of engagement in program evaluation. Low level users were characterized as simply responding to the external pull for accountability while high level users promoted an internal push for engaging in evaluation, conducting a mix of internal and external evaluations, incorporating more scientific and rigorous methods, and using evaluation in myriad ways.

Based on their qualitative analysis and comparing leaders who engaged in a high versus low level of evaluation engagement, Alaimo and Van Slyke (2006) developed a model of how leaders impact and build capacity for using evaluation not just for accountability purposes, but for learning in their organizations (Figure 10).

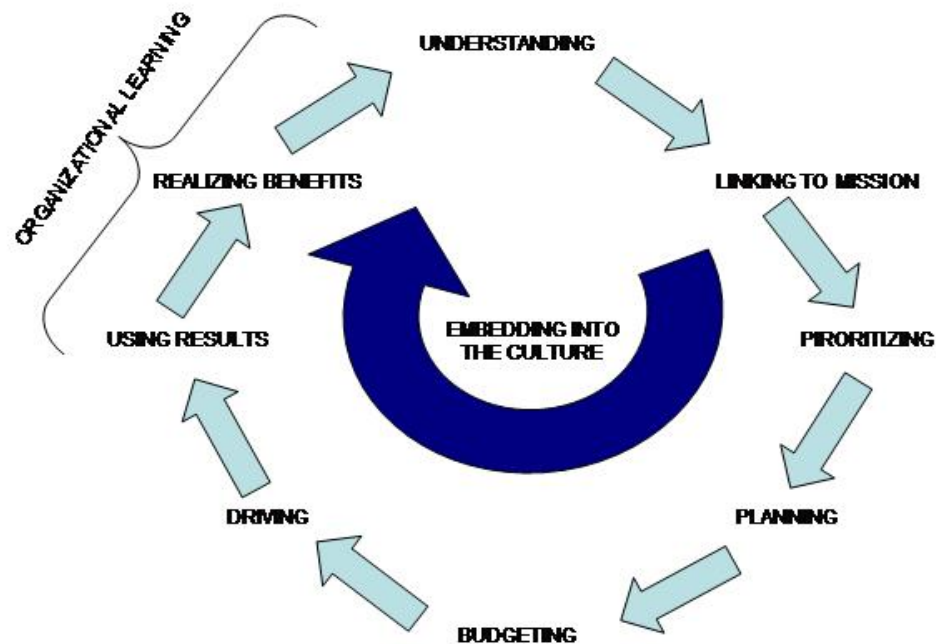


Figure 10: The role of the leader in building capacity for program evaluation (Alaimo & Van Slyke, 2006, p.9).

The authors suggested that leaders who understood the evaluation process integrated evaluation activities with mission and used evaluation results for strategic planning purposes, thereby signaling the importance of evaluation to stakeholders. This process was further supported by a self-reflective leader, who ensured there was appropriate budgeting and

support for the process. Alaimo and Van Slyke (2006) emphasized that the leader's ability to self-reflect and successfully utilize feedback individually manifested itself in the way they promoted using evaluation as an opportunity for organizational learning, leading to a culture that embeds evaluation as part of a continual learning process.

While Alaimo and Van Slyke (2006) acknowledged that their study was hampered by a small non-representative sample that focused simply on the executive director as the unit of analysis, ignoring other stakeholders, their study did provoke some insights into the leader's potential role in moving an organization beyond accountability toward organizational learning. Organizations that were depicted as using evaluation to improve their programs and organizations had leaders that believed in the usefulness of the evaluation process, used data to inform their own decision-making practices, and felt evaluation was a priority within their organization. What is not known is how the role of a leader supportive of evaluation interacts with other factors, such as the characteristics underlying organizational learning and stakeholder engagement, to facilitate using evaluation for organizational learning. Other factors that were alluded to but not analyzed in the study included issues such as how education, training, and/or previous experience with evaluation impacted leader's perceptions of the evaluation process. In addition, while this study effectively utilized qualitative data to examine the role of leadership with a small number of human service organizations, it would be useful to support this with quantitative data and look at the role of leadership evaluation characteristics on using evaluation using a larger sample of nonprofits from multiple service areas.

Summary of Empirical Research Surrounding the Factors Associated with Implementation
and Use of Evaluation for Organizational Learning: Resulting Research Questions and a
Proposed Model for the Nonprofit Sector

Evaluation capacity building is defined as the “intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine” (Stockdill, Baizerman, & Compton, 2002). A major deficiency in ability to build evaluation capacity in the nonprofit sector is the lack of empirical data on factors associated with the implementation and use of evaluation for organizational learning. The nonprofit management literature recognizes that evaluation approaches adopted by for-profit organizations hold little relevance for measuring the mission success of nonprofits, yet offer little data on nonprofit evaluation practice and effectiveness (Drucker, 1990; Forbes, 1998). While nonprofit and evaluation researchers have described evaluations at the local, national, foundational or government level, this has primarily been focused on case study approaches with descriptive or prescriptive results (Fine, Thayer, & Coghlan, 1998; McNelis & Bickel, 1996; Morley, Hatry, & Cowan, 2002; Morley, Vinson, & Hatry, 2001; Poole et al., 2000; Sawhill & Williamson, 2001; U.S. GAO, 1998). In addition, while researchers have theorized about the organizational and environmental factors that may impact the quality of evaluation in the nonprofit sector, there is practically no empirical evidence base on the importance of these factors, on how they interact, or on how they impact nonprofits’ implementation and use of evaluation to benefit programs and build nonprofit capacity (Hatry, van Houten, Plantz, & Greenway, 1997; Newcomer, 1997; Taylor & Sumariwalla, 1993).

With the understanding that empirical research on the factors associated with the implementation and use of evaluation for organizational learning in the nonprofit sector is limited, this review also drew on literature from the business, management and evaluation fields to identify some of the factors that should potentially be included in this type of research. This literature review highlighted multiple factors that could impact use of evaluation for organizational learning in the nonprofit sector, including: organizational characteristics, the presence of a learning culture, how evaluation is implemented, stakeholder engagement and supportive leadership. The relationships between these factors and use of evaluation will need to be assessed to determine which of these factors should be included in any research model.

Organizational Factors: Correlations

Theoretical research around organizational characteristics impacting the nonprofit sector emphasized that variables such as organizational size, budget, age, time doing evaluation and type of organization could impact nonprofit evaluation practices (Carman, 2005; Cook, 1988; DiMaggio, 1988; Galaskiewicz & Bielefeld, 1998; Gronjberg, 1993). Respondent characteristics, such as length of time with the nonprofit, may also impact results. In order to understand the relationship between organizational and respondent characteristics and to control the effect on future research questions, it will be critical to control for any variability related to these characteristics. This results in a set of initial research questions, which, if significant, will be used as covariates in any future research questions:

1. Is there a correlation between any of the organizational characteristics (size, age, budget, time doing evaluation and type of nonprofit) and use of evaluation by the nonprofit?

Multiple survey instruments have been developed in support of the construct of Organizational Learning, all of which emphasize the importance of organizations having the capability to learn, as defined by dimensions such as managerial commitment, systems perspective, openness and experimentation and knowledge transfer and integration. While research has been conducted looking at direct correlations between dimensions of organizational learning and organizational performance, no research has looked at the importance of organizational learning capability in the nonprofit sector and its relevance to other factors that may impact performance, particularly in regards to use of evaluation by organizations. This results in a researchable question for the nonprofit sector to assess whether the dimensions underlying organizational learning are correlated with the nonprofit's use of evaluation:

2. Do the dimensions underlying organizational learning influence use of evaluation by the nonprofit?

A small, but growing body of research has started to build a knowledge base to empirically support the linkages between stakeholder engagement and improved evaluation processes and use by organizations (Brandon, 1998; Greene, 1988a, 1988b; Turnbull, 1999). In research on the nonprofit sector, Fine, Thayer, & Coghlan (1998, 2000) and Thayer and Fine (2000) used a mixed methods approach and found that higher use of stakeholders throughout the evaluation process resulted in evaluations being perceived as more useful,

credible, and satisfying to the various stakeholders. What is unknown is how correlated stakeholder engagement is with use of evaluation.

3. Does increased stakeholder engagement result in increased use of evaluation by the nonprofit?

There is currently no quantitative data that has looked at the role of nonprofit leadership in relation to use of evaluation for organizational learning. Qualitative research by Alaimo and Van Slyke (2006) developed a model that emphasized the importance of supportive leadership to the evaluation process as nonprofit leaders play a pivotal role in setting priorities, modeling the use of data to inform decisions, and creating a supportive environment for the evaluation process to take place. While this provides a useful model, there is a need for research that will provide data to support the importance of nonprofit leaders in relation to the use of evaluation in the nonprofit sector. This would assess whether supportive leadership evaluation characteristics, such as nonprofit leaders' experience with evaluation, evaluation philosophy and perception of evaluation practice positively influence the use of evaluation by the nonprofit:

4. Do leadership evaluation characteristics predict the use of evaluation by the nonprofit?

Multivariate Relationships

Assessing the bivariate correlations between organizational characteristics, the presence of a learning culture, stakeholder engagement, leadership evaluation characteristics

and the dependent variable of use of evaluation by the nonprofit should reveal which factors need to be included in any multivariate model.

Research on evaluation use is primarily drawn from the for-profit or governmental sectors. While this research has been dominated by small case study approaches, more than 30 years of literature emphasize some overlapping factors. Stakeholder engagement in the evaluation process is critical to increasing use of results. Simply engaging stakeholder is insufficient however, as the context in which evaluation takes place impacts how the evaluation process is conducted and whether and how results are used. Consequently, if there are significant bivariate correlations between use of evaluation by the nonprofit and stakeholder engagement or organizational characteristics, such as culture and learning capacity, in addition to leadership evaluation characteristics, then these factors need to be included in any research model of factors facilitating the use of evaluation for organization learning in the nonprofit sector.

What is not known is the combined effect of the factors of stakeholder engagement, leadership evaluation characteristics and the dimensions underlying organizational learning on use of evaluation by the nonprofit. Does increased stakeholder engagement result in increased evaluation use by the nonprofit or do organizational characteristics, such as the dimensions underlying organizational learning and leadership evaluation characteristics play a larger role? If the dimensions underlying organizational learning are significant, do they continue to have an impact on evaluation use if other significant factors, such as stakeholder engagement and leadership evaluation characteristics are included in the research model?

These questions result in the need for multivariate analysis that looks at the impact of multiple factors simultaneously and asks the research question:

5. What is the combined effect of stakeholder engagement, leadership evaluation characteristics and the dimensions underlying organizational learning on the dependent variable of use of evaluation by the nonprofit?

Mediating Factors

Theoretical and case study research by Ebrahim (2005a, 2005b) and Christensen & Ebrahim (2006) emphasized that in the nonprofit sector, implementation does not necessarily result in use, especially for organizational learning. While no research has looked at how differential implementation of evaluation activities impacts use, Carman (2005) and Carman and Millesen (2004) found considerable variability in how evaluation was being both defined and implemented across nonprofit organizations, which could potentially impact how a nonprofit uses evaluation. As it is unknown whether a higher level of implementation relates in increased use of evaluation, it will be critical to assess whether there is a bivariate correlation between implementation and use of evaluation:

6. Is there a correlation between implementation and use of evaluation by the nonprofit?

If there is a relationship between implementation and evaluation use by the nonprofit, then implementation could potentially mediate any significant relationships between stakeholder engagement, leadership evaluation characteristics, the dimensions underlying organizational learning and use of evaluation by nonprofits.

Qualitative research by Taut and Alkin (2003) emphasized the importance of stakeholder engagement throughout the nonprofit evaluation process as a way to increase implementation. In addition, Carman (2005) analyzed qualitative data and found that leadership evaluation characteristics, such as experience with evaluation and positive perceptions of the benefits of evaluation to their organization are positively related to implementation of evaluation. What is not known is how the factors of stakeholder engagement and leadership evaluation characteristics interact to promote implementation of evaluation activities. Consequently, it is unknown whether implementation mediates the relationship between other factors, such as stakeholder engagement, leadership evaluation characteristics, the dimensions underlying organizational learning and the nonprofit's use of evaluation. This results in a researchable question to measure whether implementation acts as a mediator in the multivariate model:

7. Is the combined effect of stakeholder engagement, leadership evaluation characteristics, and the dimensions underlying organizational learning on the dependent variable of use of evaluation by the nonprofit mediated by implementation?

Moderating Factors

In addition to looking at the direct and mediating effects of stakeholder engagement, leadership evaluation characteristics, the dimensions underlying organizational learning and implementation on use of evaluation by nonprofits, it is also important to assess whether there are organizational characteristics, such as size and type of nonprofit, which interact to

impact the strength of any relationships. Previous research has suggested that a nonprofit's size, budget, age and/or type of organization could impact nonprofit evaluation practices (Carman, 2005; Cook, 1988; DiMaggio, 1988; Galaskiewicz & Bielefeld, 1998; Gronjberg, 1993). For example, larger nonprofits with more resources may be more likely to implement and use evaluation than smaller nonprofits. In addition, human service nonprofits potentially have more accountability requirements than environmental or advocacy nonprofits. If any interaction effects are found, this could more clearly define under what conditions these variables (stakeholder engagement, leadership evaluation characteristics, the dimensions underlying organizational learning and implementation) are associated with use of evaluation by the nonprofit. Therefore, it is important to test any moderating effects of these organizational characteristics on the correlations between stakeholder engagement, leadership evaluation characteristics, the dimensions underlying organizational learning and implementation on use of evaluation by the nonprofit:

8. Are any of the correlations (between stakeholder engagement, leadership evaluation characteristics, the dimensions underlying organizational learning and implementation) with use of evaluation by the nonprofit moderated by organizational characteristics, such as size, budget, age, funder, and/or type of organization?

Summary of Research Questions and Proposed Model

The purpose of this current research is to conduct a cross-sectional quantitative study of nonprofits' evaluation practices to gain insight into the multiple factors, particularly the organizational factors, which facilitate the use of evaluation by the nonprofit. Approaching

the integration of evaluation into nonprofits through a social innovation paradigm, this research seeks to better identify the core characteristics of both the social innovation (evaluation process) and context (the nonprofit organization) that are associated with positive outcomes. While this research is grounded in a pro-innovation bias that assumes that nonprofits can and should engage in evaluation, the goal is that evaluation can and should be a tool that benefits organizations internally by enabling nonprofits to adapt, improve and become more effective. It should be noted that this researcher understands that the transformation of evaluation from accountability to an organizational learning tool cannot simply happen at the individual nonprofit level and requires structural changes at multiple levels, both internal and external to the organization.

Based on the literature, a model of the factors associated with the implementation and use of evaluation for organizational learning is proposed (Appendix A). By looking at factors identified in the literature simultaneously, I hope to identify which factors (stakeholder engagement, leadership evaluation characteristics and the dimensions underlying organizational learning) are associated with use of evaluation for organizational learning, assess the combined impact of these factors, whether the effect of these factors on use of evaluation is mediated by implementation, and whether the strength of any of these relationships is moderated by organizational characteristics, such as size, budget, age, funder and type of nonprofit.

Contributions to Field

This ongoing research builds on a small body of current empirical research about nonprofits and program and organizational evaluation and allows some of the results in North Carolina to be compared to work in other states (Carman, 2005; Carman & Millesen, 2005; Fine et al., 1998). These data will provide an overview of current nonprofit evaluation practices and the function of evaluation in the nonprofit sector. This will help identify some potential challenges in regards to using evaluation for organizational learning in addition to understanding how to develop interventions that will build evaluation capacity in North Carolina's nonprofit sector. This information could help nonprofit leaders, funders, and support organizations develop systems that will facilitate using evaluation as an organizational learning tool.

An important research contribution is that there is currently no empirical work that has looked at the interconnections between leadership evaluation characteristics, organizational characteristics, stakeholder engagement, the dimensions underlying organizational learning and implementation and use of evaluation in the nonprofit sector. Any previous work completed around this subject area has largely been based on available theory, single case studies, or small qualitative interview studies. In addition, empirical work that has defined stakeholder engagement, implementation and use of evaluation has relied predominantly on descriptive checklists rather than conceptual rubrics that classify level of implementation, stakeholder engagement, and use of evaluation by organizations. This study should help to develop clearer, more standardized measures that are grounded in the literature of evidence-based evaluation practice.

IV: METHODOLOGY

This dissertation focuses on examining the interrelationships between organizational and leadership evaluation characteristics and looking at the factors associated with use of evaluation in nonprofits, using an existing data set. This chapter will review the research design, provide details around the research process and survey development, characterize the participants, define the measures, explain the statistical analyses that will be conducted to answer specific research questions, describe univariate statistics, and discuss the approach taken toward missing data.

Research Design

This is a cross-sectional study of the nonprofit grantees of three diverse foundations within North Carolina, a large statewide foundation based in Winston-Salem (Z. Smith Reynolds Foundation), a regional community foundation in Asheville (The Community Foundation), and a small local foundation in Raleigh (John Rex Endowment). The specific foundations included were not random, but selected due to their distribution across North Carolina, the diversity of their grantees, and their willingness to cooperate in this research and share their details and contact information for previous grantees.

Process

This comprehensive survey utilized multiple closed and open-ended questions to examine a variety of factors that impact the capacity and willingness of nonprofit organizations to conduct evaluation that facilitates organizational learning. This research

will only use the variables that help to answer the research questions, but all the variables from the survey are outlined in Appendix B.

In cooperation with the Institute for Nonprofits at North Carolina State University, The Community Foundation of Western North Carolina, the John Rex Endowment, and the Z. Smith Reynolds Foundation, data were gathered over a four week period from March 24 to April 19, 2006. The executive directors of each foundation sent a cover e-mail (example attached as Appendix C) informing the nonprofits of their support for this upcoming research and inviting their previous five years' grantees ($N=448$) to participate in an on-line survey about how to make evaluation more beneficial to their organizations. I then e-mailed out a link to an on-line survey to all of the executive directors of the foundations' past five year's grantees for which I could verify contact information.⁸ Two weeks later, the executive directors of each foundation sent out a follow-up e-mail repeating the invitation to participate, which I followed up one week later with a reference to the earlier e-mail and a clear cutoff date. This is in line with Dillman's (2000) recommendation for a four contact e-mail survey strategy, which is purported to obtain a response rate comparable to a mail survey.

Survey Development

Based on an extensive literature review, insights from formal and informal meetings with nonprofit representatives, and nonprofit academic researchers, a 26-question on-line survey utilizing both closed and open-ended questions was developed using Dreamweaver©

⁸ Each foundation sent their databases from the previous five years' grant cycles and independently of the survey, all contact information was verified via e-mail, web sites, and over the telephone. This resulted in a total sample of 448.

and posted on-line using inForm v3©. A copy of this on-line survey instrument can be found at http://www4.ncsu.edu/~dmmurphy/eed_survey_test.htm and is also attached as Appendix D. In order to refine the instrument and survey process, Dillman's (2000) principles for constructing a questionnaire and stages of pretesting were adhered to throughout the survey development process. The survey was reviewed by knowledgeable colleagues and piloted with ten nonprofit executives who were not part of the sample. These nonprofit executive directors completed the on-line survey in one location and I then conducted a focus group to elicit their critique of the process and content. This provided feedback on issues related to time, ease of use and understanding, clarity of questions, and any improvements that could be made, allowing me to refine the instrument further.

Participants

Using the National Taxonomy of Exempt Entities (NTEE) classification system,⁹ the 283 respondents represent a variety of nonprofits. In terms of the broad nonprofit categories represented; of these 283 organizations, 45 (17%) are environment related, 43 (16%) are human service, 26 (10%) are community improvement, 25 (9%) are civil rights, advocacy, and social action, 22 (8%) are education related, 14 (5%) are youth development, 13 (5%) are housing, 12 (4%) are multipurpose, and 10 (4%) are health related.¹⁰ Using annual operating budget as a measurement of size, 83 nonprofits (31%) had budgets less than \$250,000, 98

⁹ The twenty-six major groupings used by NTEE can be found on the National Center for Charitable Statistics' web site at: <http://nccsdataweb.urban.org>.

¹⁰ Fourteen organizations did not answer this. The remaining sixty organizations represent eleven different NTEE categories

nonprofits (37%) had budgets of between \$250,000 and \$1,000,000, and 86 nonprofits (32%) reported budgets of over \$1,000,000.¹¹

The overall response rate to the on-line survey was 64%, though response rates varied between the three foundations (from 52 to 75%). I attribute this relatively high overall response rate to the preceding cover letter and follow-up e-mail sent by the foundations directly to their grantees. As one nonprofit noted, one of these foundations is one of the few major funders for environmental nonprofits, which may explain the high number of environment related organizations. This means that while a wide array of nonprofits responded to the study, the sample generally has greater annual operating budgets (Table 5) and includes significantly more environmental organizations (Table 6) than is typical across North Carolina (Urban Institute, 2006). Therefore, the sample is not representative of North Carolina's nonprofits and results should be interpreted cautiously.

Table 5

Budget Comparison: Sample and North Carolina

Annual Operating Budget	Sample	NC
Less than \$250,000	31%	80%
\$250,000 - \$1,000,000	37%	11%
More than \$1,000,000	32%	9%

¹¹ Data on operating budget was not included for all organizations (missing data=17).

Table 6

NTEE Classification Comparison: Sample and North Carolina

NTEE Classification	Sample	NC
Environment	17%	2%
Human service	16%	15%
Community improvement	10%	4%
Civil rights, social action, advocacy	9%	<1%
Education	8%	19%
Youth development	5%	2%
Housing	5%	5%
Multipurpose	4%	Unknown
Health	4%	8%

Measures

The following section will describe the construction of the variables as well as descriptive univariate statistics. As can be seen from the survey (Appendix D or available on-line at: http://www4.ncsu.edu/~dmmurphy/eeed_survey_test.htm), the survey utilized a variety of response choices including checklists, open-ended spaces, five point frequency scales, seven-point Likert scales, and three-point response scales (e.g. not important, moderately important, very important). The survey included both positively and negatively worded items, which were all reverse coded prior to analysis. Appendix E shows the frequencies and percentages for all categorical variables. Descriptive statistics for all numeric variables are shown in Appendix F. The distributional characteristics (skewness and kurtosis) of numeric variables are discussed only when the distribution fails to approximate a

normal distribution, characterized as having a skewness of less than 3 and a kurtosis less than 10 (Kline, 2005). If organizations were eliminated from analyses due to their skewed influence on the results, explanations will be provided.

Organizational Characteristics: Covariates

Theoretical research around organizational and environmental characteristics impacting the nonprofit sector emphasized the importance of including organizational size, age, type of organization and time doing evaluation (Carman, 2005; Cook, 1988; DiMaggio, 1988; Galaskiewicz & Bielefeld, 1998; Gronjberg, 1993). Therefore, multiple items were included to assess whether there were differences based on these variables. These included Budget, measured as a continuous variable using annual operating budget reported (question 23) and Staff Size (addition of questions 20 and 21), Organizational Age in years (question 22), type of nonprofit using the NTEE classification system (question 18), and time in years of doing evaluation or Evaluation Years (question 5). Budget may be the strongest predictor of nonprofit size as Staff Size is often confounded by how volunteers are counted and type of nonprofit, in addition to the fact that number of paid staff is not included in public data sets, such as 990 forms.

Staff Size was positively skewed (skewness=8.2, kurtosis=79) due to the influence of five organizations with staff sizes above 1000 (mean=68). These five organizations were removed from the variable to approximate a normal distribution, (Kline, 2005), which reduced the staff size mean to 39 (Appendix F). Budget was positively skewed (skewness=11.1, kurtosis=129) due to three extreme outliers with budgets below \$500 or

above \$225,000,000. These responses were not congruent with the organization's 990 reporting and were removed from the variable to approximate a normal distribution (Kline, 2005), with a mean budget of \$178,444 (Appendix F). Evaluation Years was slightly positively skewed (skewness=4.4, kurtosis=26.3) due to three outliers with Evaluation Years of more than 90 (with the next organization being 56 years). These three organizations were removed from the variable to approximate a normal distribution (Kline, 2005), with a mean age of 21 years. These organizational variables were included as covariates in the analyses, along with the grantee's foundation, to assess any group differences and to ensure that age, size, and type of nonprofit were included if they were found to moderate any of the research results.

Dependent Variable: Use of Evaluation

Research on evaluation use is primarily drawn from the for-profit or governmental sectors. One major issue revolves around the construct of use being hampered by a lack of concrete operationalization. There are multiple definitions of use, which include cognitive (such as learning about evaluation, increased interest in evaluation) as well as behavioral uses (such as instrumental activities) that can potentially result in organizational learning. While there are diverse definitions of use of evaluation, there are no existing measures to capture this and as this research is focused on uses of evaluation by the nonprofit for organizational learning, this measure will consist of actual (behavioral) rather than perceived (cognitive) uses of evaluation. Due to the lack of existing measures, it was important to combine

theoretical frameworks with factor analysis on any measures of use to potentially identify a scale of use of evaluation in the nonprofit sector.

The dependent variable of *use of evaluation* in this research is focused on behavioral uses of evaluation and includes multiple items designed to capture how nonprofits instrumentally utilize evaluation. In order to ensure that there could be some kind of comparability with other research, these items built on the existing literature surrounding use. In particular, the multi-method research of program evaluation practice in the nonprofit sector conducted by Fine et al. (1998) and Carman (2005) was reviewed for items that could be adapted, supplemented, and used. Both authors developed these items through an iterative process that included using empirical research to develop items, supplementing these with findings from open-ended interview questions, follow-up probes, and a survey pilot process. Fine et al. (1998) and Carman (2005) then used checklists to dichotomize responses (Yes/No) for each potential use of evaluation and reported descriptive results individually.

To build on and establish a more cohesive measure of evaluation use in nonprofits, I had a series of ongoing conversations with Joanne Carman (Carman, 2005). Based on this dialogue and the feedback from the piloting process with nonprofit executive directors, I made several amendments to the use items. Carman's (2005) original items on evaluation use allowed respondents to simply check all categories without discerning levels of use, which did not allow for much range in responses. Consequently, I changed the question from a *Check all that apply...* to a question that read: *How often does your organization use evaluation for the following purposes...?* and adapted response categories from a simple yes/no or checklist to a frequency scale with five levels: never, seldom, about half the time,

usually, and always. Based on the literature and adapting items developed by Fine et al. (1998) and Carman (2005), the following items were included in the survey (Figure 11) with a (radio button) frequency scale:

8. How often does your organization use evaluation for the following purposes? a. For outreach b. For public relations c. To help make changes in existing programs and/or services d. To help make changes in organizational practices e. For strategic planning purposes f. To help us get new funding g. To make decisions about fiscal allocations h. To make decisions about staffing i. To help develop new programs and/or services j. To help us establish program goals or targets k. To help us establish organizational goals or targets l. To report to funders m. To report to the board n. To gain support for evaluation among staff or governing board

Figure 11: Survey questions related to use of evaluation.

As previously noted, one ongoing critique of the research around evaluation use is the considerable variability in how use is operationalized as a dependent variable (Cousins & Leithwood, 1986; Leviton, 2003). While some have attempted to create typologies of use (Greene, 1988a, 1988b; Leviton & Hughes, 1981), there is no empirical work around evaluation use, especially by nonprofits, that has operationalized and measured the concept of use. Any work that has looked at use generally treats items individually and is focused on descriptive results. Consequently, prior to analysis, this dissertation used exploratory factor analysis in an attempt to examine factors that make up a scale or sub-scales of use.

Responses to the fourteen-item use of evaluation question were subjected to an exploratory factor analysis using squared multiple correlations as prior communality estimates. The principal factor method was used to extract the factors, and this was followed by a promax (oblique) rotation. A scree test suggested three meaningful factors, which accounted for 99 percent of the variance in the data set. These factors could be conceptually as well as statistically supported and were labeled Accountability, Internal Learning, and Image Building.

In interpreting the rotated factor pattern, an item was said to load on a given factor if the factor loading was .40 or greater for that factor, and was less than .40 for the other. Using these criteria, one item was eliminated (8j) as it cross-loaded above .40 on two factors. Consequently, five items were found to load on the Accountability factor, four items were found to load on the Internal Learning factor, and three items were found to load on the Image Building factor. Scale reliability was assessed using Cronbach's (1951) standardized coefficient alpha and reliability estimates for each factor were above .70. Questionnaire items and corresponding factor loadings and coefficient alpha are presented in Table 7.

Table 7

Use of Evaluation Items, Factor Loadings on Each Scale, Variance Explained, and Coefficient Alphas

Use of Evaluation Items	Internal Learning	Accountability	Image Building
Changes in programs and/or services	78	-18	21
Changes in organizational practices	77	-12	15
Strategic planning	57	15	3
To develop new programs and/or services	46	28	8
To establish organizational goals or targets	73	29	-14
To secure new funding	0	45	31
To make decisions about fiscal allocations	11	62	9
To make decisions about staffing	17	64	-2
To make decisions about fiscal allocations	11	62	9
To make decisions about staffing	17	64	-2
To report to funders	-15	70	12
To report to the board	6	68	3

Table 7 (continued)

Use of Evaluation Items	Internal Learning	Accountability	Image Building
For outreach	0	7	65
For public relations	8	7	65
<i>Variance explained</i>	<i>.78</i>	<i>.11</i>	<i>.10</i>
<i>Coefficient alpha</i>	<i>.89</i>	<i>.82</i>	<i>.74</i>

Mean scale scores for these three scales of use: Internal Learning, Accountability and Image Building were calculated by summing scale scores for each nonprofit and then dividing by the number of items in the scale. All three dependent use variables approximated a normal distribution (Appendix F).

Independent Variables

The independent variables for the study included multiple measures designed to capture both observed (manifest) and underlying (latent) factors that impact use of evaluation by the nonprofit. Based on an extensive literature review, the factors highlighted as important to the use of evaluation included Leadership Evaluation Characteristics, Stakeholder Engagement and Organizational Learning Capability. This section offers a description of how each of these variables was operationalized and measured.

IV1: Leadership Evaluation Characteristics

There are no existing measures around Leadership Evaluation Characteristics as there is no quantitative research that has assessed the leader's role in moving an organization beyond accountability toward organizational learning. Research by Alaimo and Van Slyke (2006) concluded that organizations that were depicted as using evaluation for organizational learning had leaders that believed in the usefulness of the evaluation process, used data to inform their own decision-making practices, and felt evaluation was a priority within their organization. Other factors that have been alluded to include the importance of having experience with and/or training in evaluation. Consequently, based on an iterative process that included developing items based on the literature and ongoing dialogue with nonprofit scholars, piloting the survey with nonprofit executives, and conducting informal focus groups with both nonprofit executives and academics, an exploratory self-evaluation leadership (evaluation characteristics) scale was developed. It included seven items (Figure 12) with a 7-point Likert scale of strongly disagree to strongly agree:

Please respond to the following statements about yourself, using the scale strongly disagree to strongly agree

26) As a leader in this organization...

- a. I use systematically collected data to inform my decision-making
- b. I believe evaluation helps provide better programs, processes, services, and products
- c. I have experience using evaluation in this and/or other organizations
- d. I have received training on how to use evaluation
- e. Currently available information tells us what we need to know about the effectiveness of our programs and processes.
- f. It is a priority for me that our organization invests resources in evaluation
- g. I feel our organization is doing an effective job in conducting and using evaluation

Figure 12: Survey questions related to leadership evaluation characteristics.

These items or statements can be responded to by leaders in order to form a *basic* measure of Leadership Evaluation Characteristics that show how supportive the leader is of the evaluation process. Each item included can be loosely associated with the leader's *evaluation beliefs and behavior* (a, b, f), *experience with evaluation* (c and d), and *perception of current evaluation practices* (e, g).

Responses to the seven-item Leadership Evaluation Characteristics question were subjected to an exploratory factor analysis using squared multiple correlations as prior communality estimates. The principal factor method was used to extract the factors, and this was followed by a promax (oblique) rotation. A review of the scree test and eigenvalues suggested one meaningful factor, which accounted for 99 percent of the variance in the data set. All seven items loaded above .55 on this single factor and the reliability estimate, assessed using coefficient alpha, was acceptable at .82 (Cronbach, 1951). Questionnaire items, corresponding factor loadings, and coefficient alpha are presented in Table 8.

Table 8

Leadership Evaluation Characteristics' Items, Factor Loadings, Variance Explained, and Coefficient Alpha

Leadership Evaluation Characteristics' Items	Factor Loadings
Use systematically collected data to inform decisions	69
Believe evaluation helps provide better programs, processes, services, and products	55
Have experience using evaluation in this and/or other organizations	74
Have received training on how to use evaluation	68
Believe currently available information is effective	56
Feel evaluation is a priority for resources	62
Feel nonprofit doing effective job of conducting and using evaluation	61
<i>Variance explained</i>	.99
<i>Coefficient alpha</i>	.82

Mean scale scores for this Leadership Evaluation Characteristics' scale were calculated by summing scale scores for each nonprofit and then dividing by the seven items

in the scale. The Leadership Evaluation Characteristics' variable approximated a normal distribution (Appendix F).

IV2: Stakeholder Engagement

Very little research addresses the role of Stakeholder Engagement in the implementation and use of evaluation in the nonprofit sector. Fine et al. (2001) conducted the only research to date that looked at which stakeholders were engaged in different parts of the evaluation process. Fine et al. (2001) defined four different categories of stakeholders (staff, board members, funders and program participants) and then asked which stakeholders were involved in four different stages of the evaluation process: designing the evaluation, collecting data, analyzing and interpreting the results, and/or writing the evaluation report. Fine et al. (2001) then defined levels of Stakeholder Engagement by assigning organizations one point for every stakeholder that was engaged in the four different stages of the evaluation process; this resulted in a maximum of 16 points. After looking at descriptive results, Fine et al. then created three evenly sampled groups (i.e. split the organizations into thirds) and defined the organizations as having low (0-2), medium (3-5), and high stakeholder involvement (6-11). While this is an understandable way to group organizations, it is not based on any theoretical premise of what constitutes a low, medium, or high level of stakeholder involvement.

This research has emphasized that *who* is involved in the different stages of the evaluation process makes a difference in whether the nonprofit uses evaluation for organizational learning. In an attempt to replicate some of these findings, but improve upon

the existing measures, I reviewed the literature and engaged nonprofit academics and executive staff at several foundations and nonprofits in an ongoing dialogue to articulate the different steps of the evaluation process and the different stakeholders that could potentially be involved. Based on these discussions and the impetus to include Greene's (1988a; 1988b) notion of *meaningful* stakeholder participation, as defined by shared decision making, I added four extra evaluation process steps: decision making regarding evaluation planning, judging the appropriateness and quality of the evaluation's methodology, reading and using evaluation results, and deciding how evaluation results are used by the organization. In addition, to gain more information about not only which stakeholder groups were involved (staff, board members, funders, program participants) but how they differentially participated, I added the category of external evaluator and expanded on the staff group to differentiate between internal, executive management staff, internal and other evaluation staff, and students or volunteers (Figure 13).

7. Please indicate all the stakeholders who participate in the following steps of the evaluation process: (please check ALL that apply): *If you do not do one of these activities, please check not applicable (N/A)*

Internal, executive/management staff
 Internal, evaluation staff
 Other internal staff
 Board members or board committees
 Students/volunteers (i.e., interns, class projects)
 External evaluator
 External agency or funder
 Program participants
 Other (please specify)
 N/A We do not do this

- a. Decision-making regarding evaluation planning
- b. Design of the evaluation
- c. Judging the appropriateness and quality of the evaluation's methods
- d. Collecting data about programs or practices
- e. Interpreting the meaning of the evaluation data
- f. Writing the report
- g. Reading and using evaluation results
- g. Deciding how evaluation results are used by the organization

Figure 13: Survey questions related to stakeholder participation.

To look more closely at who was engaged in which stages of the evaluation process, eight different stakeholders were assessed: executive/management staff, internal evaluation staff, other internal staff, students/volunteers, board members, funders (external agencies), and program participants. Building on the research that suggested that stakeholder group matters and that meaningful stakeholder involvement is important, this measure of

Stakeholder Engagement was based on the assumption that these eight different stakeholders would represent different dimensions of Stakeholder Engagement.

To define a clear measure of Stakeholder Engagement, the dichotomous responses to the items were explored using a heuristic method based on factor analysis of the matrix of tetrachoric correlations for all item pairs (Uebersax, 2000). The tetrachoric correlation between two dichotomous items estimates the Pearson correlation one would obtain if the two constructs were measured continuously (Drasgow, 1988). Once the tetrachoric matrix was developed, it was subjected to an exploratory factor analysis using squared multiple correlations as prior communality estimates.

Preliminary analysis revealed a lack of variability across a number of stakeholders, including participants, funders, students, and volunteers. 75% of nonprofits indicated that no participants engaged in any stage of the evaluation process. Similar statistics were seen across other stakeholder groupings, including funders (70%), students (74%) and volunteers (72%). As there was insufficient data for ongoing analyses, these stakeholders were excluded from the proposed exploratory factor analysis. Instead, four groupings were included for the Stakeholder Engagement measure: Executive Director, Board Member, Internal Evaluator, and Staff Engagement.

For each factor analysis, the principal factor method was used to extract the factors, and this was followed by a promax (oblique) rotation. A review of the scree test and eigenvalues suggested one meaningful factor for each stakeholder. All eight items loaded above .70 on one single factor and accounted for more than 80 percent of the variance in each data set. Questionnaire items and corresponding factor loadings are presented in Table 9.

Table 9

Stakeholder Engagement Item, Groups, Factor Loadings, and Variance Explained

Stakeholder Engagement item	Executive Director	Internal Evaluator	Staff Engagement	Board Member
Evaluation planning	95	98	79	85
Evaluation design	92	95	95	78
Methodology	91	95	97	80
Collecting data	78	94	81	85
Interpreting meaning	91	96	93	89
Writing report	87	96	99	80
Reading and using results	95	99	95	83
Deciding how results are used	94	92	81	87
<i>Variance explained</i>	<i>.86</i>	<i>.85</i>	<i>.85</i>	<i>.81</i>

In line with McGrath (2001), to ensure these groups had discriminant validity, I assessed the correlations between each of these stakeholders. While the correlations between the dimensions of Executive Director and Staff Engagement and Executive Director and Board Members were significant ($p < .01$), these correlations were minimal ($< .21$) and I therefore concluded that the stakeholders were sufficiently differentiated to be treated independently. As these discriminant and factor analyses provide sufficient support for using these four rationally created groupings and to ensure results are easily interpretable, I used manifest continuous measures for each stakeholder grouping. These four measures:

Executive Director, Internal Evaluator, Staff Engagement, and Board Member have a score range of 0-8. As can be seen from the descriptive statistics in Appendix F, all four Stakeholder Engagement variables approximated a normal distribution (Kline, 2005).

IV3: Organizational Learning Capability

To measure the capability of an organization to use evaluation for organizational learning, I reviewed the multiple scales that have previously been developed in an attempt to find a usable scale that could be incorporated into a survey looking at multiple variables simultaneously. While the Dimensions of the Learning Organization Questionnaire (DLOQ) (1993) has the most empirical data to support it (primarily conducted by the authors), it consists of 43 items. In addition, its focus on the for-profit sector, measures of financial performance and larger resource-extensive businesses makes it a somewhat inappropriate tool for the nonprofit sector. The Readiness for Organizational Learning and Evaluation (ROLE) instrument (Preskill and Torres, 1999), in addition to being a 78-item questionnaire designed to be administered to multiple people within a single organization and used as a benchmark, assumes that organizations are engaged in evaluation processes. Whether nonprofits are engaged in evaluation is a researchable question for the sector, and, as Preskill and Torres (1999) emphasized that ROLE should be implemented in its complete form, ROLE was also deemed inappropriate for use in the nonprofit sector.

The Organizational Learning Capability scale.

Jerez-Gómez, Céspedes-Lorente, & Valle-Cabrera (2004) developed a shorter measurement tool to identify the underlying characteristics of organizational learning or the

capability of an organization to learn systematically. Based on an extensive review of the dimensions of Organizational Learning and validated scales developed by Goh & Richards (1997), Hult & Ferrell (1997), and Oswald, Mossholder, & Harris (1994), Jerez-Gómez et al. (2004) created a multidimensional Organizational Learning Capability Scale consisting of 16-items that focuses on managerial commitment, systems perspective, openness and experimentation, and knowledge transfer and integration as the key elements that construct organizational learning capability.¹² Jerez-Gómez et al. (2004) perceive organizational learning as a latent multidimensional construct that underlies these four dimensions (managerial commitment, systems perspective, openness and experimentation, and knowledge transfer and integration). In addition, Jerez-Gómez et al. (2004) noted that these four dimensions have shared characteristics and emphasized the relevance and inclusion of stakeholder participation, in terms of employee participation, teamwork, cooperation, and involvement, within each dimension.

To check the scale's validity, in terms of content validity, reliability, convergent validity, and discriminant validity, Jerez-Gómez et al. used a sample of managers from 111 chemical companies. The authors initially used principal components and confirmatory factor analysis to corroborate the existence of the four dimensions, which explained 63 percent of the total variance. They further conducted second order factor analysis using the Generalized Least Squares (GLS) method, which provided empirical backing to the proposed organizational learning structure where learning was proposed as a latent construct. Jerez-

¹² This is synonymous with the construct of the dimensions of organizational learning. The authors simply wish to make explicit that organizational learning is a latent construct that results from the underlying characteristics of these dimensions.

Gómez et al. noted the limitations of the study in terms of response bias, small sample size, and lack of already-validated scales. Despite this, the comprehensive statistical analysis employed conjoined with the comparative shortness of the scale make this scale very usable, especially for inclusion in a study that seeks to look at organizational learning as one of multiple variables. One further advantage is that the questions in this scale are not focused on aspects of performance as related to financial measures and could be easily adapted for use in the nonprofit sector.

This research adapted Jerez-Gómez et al.'s (2004) multidimensional Organizational Learning Capability scale consisting of 16 items for use in the nonprofit sector (questions 10 to 13) using the same 7-point Likert scale (strongly agree to strongly disagree). This scale was then reviewed by several nonprofit directors to ensure the wording of the questions was appropriate prior to piloting the survey in its entire form with nonprofit executives. These sixteen items correspond with four dimensions deemed critical to predict capability for organizational learning: managerial commitment, systems perceptive, openness and experimentation, and knowledge transfer and integration. Figure 14 shows the questions associated with each dimension.

--Managerial commitment (MC) (10a-e)

10a. The organization's leadership frequently involves staff in important decision-making processes.

10b. Employee learning is considered more of an expense than an investment.

10c. The organization's leadership looks favorably on carrying out changes in any area to adapt to and/or keep ahead of new environmental situations.

10d. Employee learning capability is considered a key factor in this organization.

10e. In this organization, innovative ideas that work are rewarded.

--Systems perspective (SP) (11a-c)

11a. All employees have generalized knowledge regarding this organization's objectives.

11b. All parts that make up this organization (departments, board, program staff, executive staff, volunteers, and other individuals) are well aware of how they contribute to achieving the overall objectives.

11c. All parts that make up this organization are interconnected, working together in a coordinated fashion.

--Openness and experimentation (EX) (12a-d)

12a. This organization promotes experimentation and innovation as a way of improving the work processes.

12b. This organization follows up what other organizations in the nonprofit sector are doing, adopting those practices and techniques it believes to be useful and interesting.

12c. Experiences and ideas provided by external sources (funders, technical assistance organizations, consultants, etc.) are considered a useful instrument for this organization's learning.

12d. Part of this organization's culture is that employees can express their opinions and make suggestions regarding the procedures and methods in place for carrying out tasks.

--Knowledge transfer and integration (TR) (13a-d)

13a. Errors and failures are always discussed and analyzed in this organization, on all levels.

13b. Employees have the chance to talk among themselves about new ideas, programs, and activities that might be of use to the organization.

13c. In this organization, teamwork is not the usual way to work.

13d. The organization has instruments (manuals, databases, files, organizational routines, etc.) that allow what has been learnt in past situations to remain useful, although the employees are no longer the same.

Figure 14: Survey questions related to Organizational Learning Capability.

The factor structure of the Organizational Learning Capability Scale was evaluated using confirmatory factor analysis in SAS (Hatcher, 1994; Jöreskog & Sörbom, 1989). Jerez-Gómez et al. (2004) specified a four dimensional model that predicted the second order factor of organizational learning capability. A second order factor structure proposed that organizational learning capability is a latent construct that underlies four dimensions, which are also latent and measured using sixteen different items. The second-order factor analysis did not provide empirical backing to the proposed organizational learning structure model. The fit of this model was inadequate with a significant Chi-square estimate: $X^2(100, 262) = 197, p < .0001$, (GFI=.89, RMSEA=.08, RMR=.11, NFI=.40, NNFI=.35, CFI=.48, PGFI=.72).

Consequently, a confirmatory modeling approach was used to focus specifically on the four-factor solution. All of the four factors (Managerial Commitment, Systems Perspective, Openness and Experimentation, Knowledge Transfer and Integration) were comprised of at least three composite variables. The model predicted by Jerez-Gómez et al. (2004) was inadequate with a significant Chi-square estimate: $X^2(99, 262) = 275, p < .0001$, (GFI=.88, RMSEA=.08, RMR=.07, NFI=.81, NNFI=.84, CFI=.87, PGFI=.73). While changes to the model could have been made based on the modification indices, Hair et al. (1999) cautioned against this action unless there is sound theoretical support for doing so. Instead, one item (10br) was removed due to the standardized loading on to its factor being less than .40 (.37) and one further item (13b) was removed due to significant loadings on all four factors. In the revised model, despite a significant Chi-square, the fit indices improved sufficiently to be considered adequate: $X^2(72, 262) = 191, p < .0001$, (GFI=.91, RMSEA=.08,

RMR=.07, NFI=.85, NNFI=.88, CFI=.90, PGFI=.72). Table 10 provides a summary of the parameter estimates for this four-factor solution.

Table 10

Organizational Learning Capability Items, Factor Loadings, and Unique Variance

Organizational Learning Capability Items	Managerial Commitment	Systems Perspective	Openness and Experimentation	Knowledge Transfer and Integration	Unique Variance
The organization's leadership frequently involves staff in important decision-making processes	.53				.28
The organization's leadership looks favorably on carrying out change in order to adapt to and/or keep ahead of new environmental situations	.62				.39
Employee learning capability is considered a key factor in this organization	.61				.37
In this organization, innovative ideas that work are rewarded	.56				.31
All employees have generalized knowledge regarding this organization's objectives		.64			.41
All parts that make up this organization (departments, board, program staff, executive staff, volunteers, and other individuals) are well aware of how they contribute to achieving the overall objectives		.75			.63
All parts that make up this organization are interconnected, working together in a coordinated fashion		.81			.64
This organization promotes experimentation and innovation as a way of improving the work processes			.69		.47
This organization follows up on what other organizations in the nonprofit sector are doing, adopting those practices and techniques it believes to be useful and interesting			.62		.39
Experiences and ideas provided by external sources (funders, technical assistance organizations, consultants, etc.) are considered a useful instrument for this organization's learning			.54		.30

Table 10 (continued)

Organizational Learning Capability Items	Unique Variance	Knowledge Transfer and Integration	Openness and Experimentation	Systems Perspective	Managerial Commitment
Part of this organization's culture is that employees can express their opinions and make suggestions regarding the procedures and methods in place for carrying out tasks	.50		.71		
Errors and failures are always discussed and analyzed in this organization, on all levels	.37	.61			
In this organization, teamwork is not the usual way to work	.24	.49			
The organization has instruments (manuals, databases, files, organizational routines, etc.) that allow what has been learned in past situations to remain useful, although the employees may no longer be the same	.23	.49			

Mean scale scores for the four Organizational Learning Capability scales were calculated by summing scale scores for each nonprofit and then dividing by the items in the scale. As can be seen from the descriptive statistics in Appendix F, all scales approximated a normal distribution (Kline, 2005).

Mediating factor: Implementation of evaluation.

Research around implementation of evaluation emphasizes that it is important to know what is being implemented in addition to the prevalence of evaluation practices across the organization. While there are no empirical measures of implementation and activities have primarily been framed in concrete and behavioral terms, then described with frequencies attached, it would be useful to have a basic level of implementation that captures what organizations are doing in addition to how prevalent these activities are across the

nonprofit. This would help distinguish nonprofits that are engaging simply in accountability practices, such as those required by a funder, from those that are moving toward organization-wide outcome measurement systems.

As a result, two questions from the survey were used to identify type of Evaluation Activities (Figure 15) and Evaluation Prevalence across organization (Figure 16) in order to develop measures of evaluation implementation. Evaluation Activities has dichotomous responses (Yes/No) as nonprofits are asked to check which activities their organization is currently practicing. Evaluation Prevalence asks how leaders would characterize their organization's overall approach to evaluation. This is divided into four separate response categories that depict whether the nonprofit evaluates no programs and organizational activities, a few programs and organizational activities, more than half, or all/almost all programs and organizational activities.

1. Does your organization currently practice any of the following management, oversight, or evaluation activities? (Please check ALL that apply to any of your organization's programs or services)

Conduct financial audits of your books
 Experience site visits by funders or regulatory agencies
 Acquire official licenses to operate programs (i.e., licenses for group homes, treatment programs)
 Participate in accreditation processes (i.e., Council on Accreditation, Council on Quality & Leadership, Joint Committee on Accreditation of Healthcare Organizations)

Conduct performance reviews and evaluations of program staff
 Review program documentation (i.e., participant or client records, case notes)
 Conduct firsthand observations of program activities
 Conduct formal program evaluations of your programs

Establish performance targets or program goals (i.e., serve 1000 people, 80% complete training)
 Monitor program implementation (i.e., to make sure programs are being delivered as intended)
 Assess whether you are meeting program goals and objectives
 Assess whether you are meeting organizational goals and objectives
 Engage in formal strategic planning processes

Produce annual reports
 Produce reports for funders about program activities
 Produce reports for funders about financial expenditures
 Produce reports for the board of directors

Use a performance measurement system (i.e., the United Way's outcome measurement system)
 Design program "logic models" (i.e., create diagrams linking inputs, processes & outcomes)
 Use a "balanced scorecard" management system (i.e., the system created by Kaplan & Norton)
 Use other management, evaluation and monitoring tools (please specify in the space provided)

Figure 15: Survey questions related to Evaluation Activities.

4. How would you characterize your organization's current overall approach to evaluation? (Please check only one statement)
We do not evaluate any of our programs and organizational activities We evaluate a few selected programs and organizational activities We evaluate more than half of our programs and organizational activities We evaluate all/almost all of our programs and organizational activities

Figure 16: Survey questions related to Evaluation Prevalence.

While there were commonalities between the Evaluation Activities' items, there was no apriori way to rationally group activities. As a consequence, to define a clear measure of implementation of Evaluation Activities, the dichotomous responses to the items were explored using a heuristic method based on factor analysis of the matrix of tetrachoric correlations for all item pairs (Uebersax, 2000). The tetrachoric correlation between two dichotomous items estimates the Pearson correlation one would obtain if the two constructs were measured continuously (Drasgow, 1988). Once the tetrachoric matrix was developed, it was subjected to an exploratory factor analysis using squared multiple correlations as prior communality estimates.

The principal factor method was used to extract the factors, and this was followed by a promax (oblique) rotation. A review of the scree test (Cattell, 1966) and Kaiser criterion (Kaiser, 1960) suggested there were seven factors with eigenvalues above 1.0, but that these seven factors accounted for only 78% of the variance in the data set. Fifteen factors were needed to account for 99% of the variance within the 21-item measure. The first factor accounted for 36% of the variance and had an eigenvalue of 7.6, while the remaining six factors explained between 5 and 9% of the variance and had eigenvalues of between 1.0 and 1.9. More importantly, applying an interpretability criterion to this measure revealed three

issues: these seven factors did not have three items that loaded on to each factor; the items that loaded on to a given factor did not share a conceptual meaning or measure different constructs; the factors did not demonstrate simple structure, in that 12 of the items loaded on to more than one factor above .40. In addition, all of the 21 items loaded above .40 on one factor. Questionnaire items and corresponding factor loadings are presented in Table 11.

Table 11

Implementation of Evaluation Activities, Factor Loadings, Eigenvalues, and Variance Explained

Activity	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Site visits	*54	35	37	31	7	-2	31
Licensing	*45	*58	33	18	*-46	28	-2
Accreditation	*66	38	20	-14	-29	-29	12
Staff reviews	*64	18	-26	-34	-3	-20	23
Board reviews	*55	10	-17	-31	36	22	*42
Document review	*62	27	-12	28	-13	*-46	-6
Observation	*42	-14	-28	*53	-13	-33	21
Formal evaluation	*64	-13	-35	13	* 45	-2	10
Establish goals	*75	-5	-6	16	-26	31	4
Monitor implementation	*75	-10	-27	4	-3	7	*-48
Assess program goals	*71	-27	*-50	1	-23	0	-9
Assess organizational goals	*64	-22	-19	-19	*-43	*42	8
Strategic planning	*53	-29	17	* -57	-20	1	6
Annual reports	*42	30	35	-17	26	16	-35
Funder reports (programs)	*70	*-47	30	30	7	-1	-11
Funder reports (finance)	*65	-40	*46	14	4	-3	-13
Board reports	*58	-32	39	-12	17	-13	38
Performance system	*65	23	-10	-18	11	-3	-9
Logic models	*63	7	10	-37	18	-40	-36
Balanced scorecard	*41	*59	-28	13	35	23	-5
Evaluation tools	*54	-14	11	28	35	32	-3
<i>Eigenvalues</i>	7.59	1.96	1.67	1.56	1.40	1.21	1.09
<i>Variance explained</i>	.36	.09	.08	.07	.07	.06	.05

As there was limited empirical support to create factors that rationally depicted different levels of implementation of Evaluation Activities, I reduced the activities to retain

as much information as possible from the nonprofits, while creating a more refined measure of Evaluation Activities. I eliminated activities that had little variability in addition to some of the more passive activities more related to management than evaluation. Variables that were excluded were site visits (q1a), licensure (q1b), accreditation (q1c), staff performance reviews (q1d), observation (q1g), assess program goals (q1k), assess organizational goals (q1l), funder reports (programs) (q1o), funder reports (financial) (q1p), and board reports (q1q). This resulted in a manifest continuous measure of the number of activities in which nonprofits' engaged, which ranged from 0-11 for the implementation variable of Evaluation Activities (Table 12). Appendix F shows the descriptive statistics for the Evaluation Activities' variable, which approximated a normal distribution (Kline, 2005).

Table 12

Evaluation Activities Measure

Activity	Definition
Board reviews	Conduct performance reviews and evaluations of board
Document review	Review program documentation (e.g., participant or client records, case notes)
Observation	Conduct firsthand observations of program activities
Formal evaluation	Conduct formal program evaluations of your programs
Establish goals	Establish performance targets or program goals (e.g., serve 1000 people, 80% complete training)
Monitor implementation	Monitor program implementation (e.g., to make sure programs are being delivered as intended)
Strategic planning	Engage in formal strategic planning processes
Annual reports	Produce annual reports
Performance system	Use a performance measurement system (e.g., the United Way's outcome measurement system)
Logic models	Design program "logic models" (e.g., create diagrams linking inputs, processes & outcomes)
Balanced scorecard	Use a "balanced scorecard" management system (e.g., the system created by Kaplan & Norton)
Evaluation tools	Use other management, evaluation and monitoring tools (please specify in the space provided)

As Evaluation Prevalence or extensiveness of evaluation activities (Figure 16) was a single, manifest measure, this was used to separate nonprofits into four groups in order to assess whether there was a bivariate correlation between prevalence of activities and the

three dependent variables, Internal Learning, Accountability, and Use for Image Building. As Evaluation Prevalence was significantly correlated with Internal Learning ($X^2=.51$, $p<.0001$), Accountability ($X^2=.41$, $p<.0001$), and Use for Image Building ($X^2=.37$, $p<.0001$), then Evaluation Prevalence was included as a second measure that predicts implementation. Appendix E shows the frequency and percentage for each of the Evaluation Prevalence groups.

Analyses

Correlations

In order to trim non-significant variables and derive a more parsimonious model, research questions 1, 2, 3, 4 and 6 were tested using the Spearman and/or Pearson correlation coefficient to assess whether there were significant correlations between the nonprofit characteristics (Staff Size, Budget, Age, Time doing evaluation, Type of Nonprofit), the four measures of Organizational Learning Capability, Stakeholder Engagement, Leadership Evaluation Characteristics, Implementation and the three dependent variables of Use of Evaluation by the nonprofit. If bivariate correlations were not significant for any variable, it was deemed appropriate to exclude these variables from further analysis.

Multivariate Relationships

To continue this trimming process at the multivariate level, multiple regression was used to answer research question five and assess whether there was a significant relationship between the three dependent variables of Use of Evaluation by the nonprofit, and the multiple significant independent variables: the four measures of Organizational Learning Capability, Stakeholder Engagement, Leadership Evaluation Characteristics, when taken as a group. Multiple regression analyses will help clarify whether each of these independent variables accounts for a statistically significant amount of variance in the variables of Use of Evaluation by the nonprofit, beyond the variance accounted for by the other independent variables.

Moderators

Research question eight assesses whether nonprofit characteristics such as Staff Size, Budget, Age, Funder, and/or Type of Nonprofit moderate any of the significant correlations between the four measures of Organizational Learning Capability, Stakeholder Engagement, Leadership Evaluation Characteristics, Implementation and the three dependent variables of Use of Evaluation by the nonprofit. This will test whether the effect of these variables on use of evaluation by the nonprofit changes at different levels of the nonprofit characteristics. One example of this would be that human service nonprofits may show a stronger relationship between Leadership Evaluation Characteristics and use of evaluation by nonprofits when compared to other types of nonprofits.

Frazier, Tix and Barron (2004) stated that multiple regression can be used to examine moderator effects when the predictor or moderator variables are categorical (Funder, Type of

Nonprofit) or continuous. Consequently, hierarchical multiple regression will be used to examine any moderating effects of these nonprofit characteristics and see which nonprofit characteristics need to be retained as moderators in the overall model.

Overall Model

Once the correlational and multivariate analyses have established which variables are significant, research question seven can be tested using an overall model. Research question seven tests whether the combined effect of the (significant) independent variables on the three dependent variables of use of evaluation by the nonprofit is mediated by the two measures of Implementation: Evaluation Activities and Evaluation Prevalence. As the multiple factors included in this research are a combination of manifest and latent variables, the most appropriate analysis to conduct is path analysis rather than latent variable structural equation modeling (SEM). Path analysis is better than regression as it can test all variables (regardless of whether they are latent and manifest) simultaneously and thus can also include Organizational Learning Capability (or any other variable) as a latent variable (or a hypothetical construct). The analyses will be conducted using PROC CALIS, a SAS procedure that can be used for path analysis, confirmatory factor analysis, SEM and other purposes.

The model that will be tested is included as Appendix A. This implies that there may be covariance between the three independent variables: Leadership Evaluation Characteristics, Stakeholder Engagement, and Organizational Learning Capability and that the three dependent variables of Use of Evaluation may be mediated by the two precursors of Implementation of Evaluation: Evaluation Activities and Evaluation Prevalence .

Dealing with Missing Data

While it is rare for any researcher not to encounter some form of missing data, there are a number of approaches that can be taken to minimize the impact of missing data and to ensure that any patterns in the missing data are accounted for in the research (Hair, Anderson, Tatham, & Black, 1999). Approaches for dealing with missing data can be categorized into four groups: use of observations with complete data only, deleting the offending case, using all of the information available as an imputation technique, or using the prevalent practice of replacing the missing item with the person mean within the scale or item mean from the sample (Egan, Yang, & Bartlett, 2004; Yang, 2003). If the missing data are considered minimal, are deemed to be missing at random, and do not significantly impact the sample size, then listwise deletion (complete case analysis) is the most appropriate technique (Allison, 2001; Little & Rubin, 1987). Listwise deletion (complete case analysis) assumes that data are missing at random and while this does decrease the sample size available, it has the advantage of providing unbiased parameter estimates.

Analysis of missing data across all the variables revealed that fifteen cases ($N=283$) had the majority of data missing on two or more variables and a further ten cases were missing the majority of data on one variable. These twenty-five organizations represent less than ten percent of the sample and data were missing at random as these organizations varied in terms of size, age, and type of nonprofit. Consequently, their exclusion from the analysis should not significantly bias the results. Fourteen other organizations had minimal data missing on one or more variables. When organizations had minimal data missing, characterized as no more than a third of the values for each variable (i.e. one item on a three-

item scale), then an expectation-maximization (EM) imputation technique was used. EM imputation is an iterative procedure that involves replacing an incomplete item with complete information based on a maximum likelihood estimate of the true value of the unobserved item (Truxillo, 2005). This is an improvement on mean substitution or imputation of missing values, which is critiqued as it reduces the variance of the variable and can negatively bias any correlations between variables (Allison, 2001; Truxillo, 2005).

V: RESULTS

This dissertation seeks to identify which factors (Stakeholder Engagement, Leadership Evaluation Characteristics and the four measures of Organizational Learning Capability) are associated with Use of Evaluation by the nonprofit. The combined impact of these factors is assessed, in addition to whether the effect of these factors on Use of Evaluation is mediated by implementation, or the strength of any of these relationships is moderated by organizational characteristics, such as Staff Size, Budget, Age, and Time Doing Evaluation. While the previous chapter discussed univariate statistics, this chapter describes the bivariate, multivariate, and path analysis findings related to each of the research questions. This dissertation used a trimming approach toward building a more parsimonious statistical model and as such, correlations were used to assess which variables to retain in the overall model.

Research Question 1: Is there a Correlation between any of the Organizational Characteristics (Size, Budget, Age, Time doing Evaluation and Type of Nonprofit) and Use of Evaluation by the Nonprofit?

To assess which organizational characteristics to retain as covariates, bivariate correlational analyses were used to derive a more parsimonious model. As the funder and nonprofit type groups had unequal numbers, Tukey's "honestly significant difference" (HSD) tests were used to assess whether groups were significantly different from each other. These analyses revealed no significant differences between Type of Nonprofit (NTEE) and the three dependent measures of Use of Evaluation: Internal Learning ($F(8,249) = 0.42; p=.91$), Publicity ($F(8,249) = 1.54; p=.14$), or Accountability ($F(8,249) =$

0.79; $p=.62$). No significant differences were found between Funder and Internal Learning ($F(2,255) = 0.33$; $p=.72$), Publicity ($F(2,255) = .83$; $p=.44$), or Accountability ($F(2,255) = 0.28$; $p=.75$). Pearson correlations were used to assess the direct relationships between Staff Size, Budget, Age of nonprofit, Time Doing Evaluation and the three dependent variables of Use of Evaluation. None of these correlations were significant as all the organizational characteristic variables correlated with the three dependent variables at .10 or below. Based on these analyses, I concluded that there were no significant correlations between any of the organizational characteristics and Use of Evaluation by the nonprofit and therefore none of these variables were included as covariates in future research questions.

Research Question 2: Do the Dimensions Underlying Organizational Learning Influence Use of Evaluation by the Nonprofit?

Pearson correlations were used to assess the direct relationships between the measures of Organizational Learning Capability: Managerial Commitment, Systems Perspective, Openness and Experimentation, Knowledge Transfer and Integration, and the three dependent variables of Use of Evaluation. As can be seen from Table 13, all of the dimensions underlying Organizational Learning Capability were significantly and positively correlated with Internal Learning, Accountability, and Image Building. Consequently, it can be determined that the dimensions underlying Organizational Learning Capability are positively associated with the Use of Evaluation by the nonprofit and these variables will be retained for the multivariate analyses.

Research Question 3: Does Increased Stakeholder Engagement Result in Increased Use of Evaluation by the Nonprofit?

Pearson correlations were used to assess the direct relationships between the four measures of Stakeholder Engagement: Board, Internal Evaluator, Internal Staff, and Executive Director and the three dependent variables of use of evaluation. As can be seen from Table 13, Board Engagement was significantly and positively correlated with Internal Learning and Accountability, but not with Image Building. Internal Evaluator Engagement was significantly and positively correlated with Internal Learning and Image Building, but not with Accountability. Internal Staff Engagement was significantly and positively correlated with Internal Learning, but not with Image Building and Accountability. Executive Director Engagement was not significantly correlated with any of the dependent variables. Consequently, it can be determined that increased Board, Internal Evaluator, and Internal Staff Engagement are positively associated with increased Use of Evaluation and will be retained for the multivariate analysis, while Executive Director Engagement will be dropped from further analysis.

Table 13

Correlations between the Predictor Variables and Evaluation Use Measures

Variable	Pearson Correlations		
	Internal Learning	Accountability	Image Building
Leadership Evaluation Characteristics	.46**	.43**	.39**
<i>Organizational Learning</i>			
Managerial Commitment	.45**	.27**	.23**
Systems Perspective	.40**	.27**	.29**
Openness and Experimentation	.43**	.29**	.25**
Knowledge Transfer and Integration	.43**	.37**	.37**
<i>Stakeholder Engagement</i>			
Board Engagement	.21**	.13*	.03
Internal Evaluator Engagement	.14*	.08	.18*
Internal Staff Engagement	.17*	.06	.02
Executive Director Engagement	.10	.06	.04
<i>Implementation</i>			
Evaluation Activities	.46**	.39**	.40**
Evaluation Prevalence	.50**	.40**	.37**

Note. * $p \leq .05$

** $p \leq .001$

Research Question 4: Do Leadership Evaluation Characteristics Predict the Use of Evaluation by the Nonprofit?

Pearson correlations were used to assess the direct relationship between Leadership Evaluation Characteristics and the three dependent variables of Use of Evaluation. As can be seen from Table 13, Leadership Evaluation Characteristics were positively correlated with Internal Learning, Accountability, and Image Building. Consequently, the Leadership Evaluation Characteristics' variable can be seen as positively predicting the three measures of Use of Evaluation by the nonprofit and will be retained for the multivariate analyses.

Research Question 5: What Is the Combined Effect of Stakeholder Engagement, Leadership Evaluation Characteristics and the Dimensions Underlying Organizational Learning on the Dependent Variable of Use of Evaluation by the Nonprofit?

Separate multiple regressions were used to assess the combined effect of the significant factors of Stakeholder Engagement, Leadership Evaluation Characteristics and the four Organizational Learning Capability measures on the three measures of Use of Evaluation: Internal Learning, Accountability, and Image Building. Using multiple regression, Internal Learning, Accountability, and Image Building scores were regressed on the linear combination of Internal Evaluator Engagement, Internal Staff Engagement, Board Engagement, Leadership Evaluation Characteristics, and the four Organizational Learning Capability measures.

The equation containing these eight variables accounted for 35% of the variance in Internal Learning, ($F(8, 249) = 16.55; p < .001$). Beta weights (standardized multiple

regression coefficients) and uniqueness indices were then reviewed to assess the relative importance of the variables in the prediction of use of evaluation for Internal Learning.¹³

Beta weights (β) and uniqueness indices (r) are presented in Table 14.

Table 14

Factors Impacting Internal Learning

Predictor	Internal Learning	
	β	r
Leadership Evaluation Characteristics	.23**	.05
Managerial Commitment	.05	<.01
Systems Perspective	.04	.02
Openness and Experimentation	.08	<.01
Knowledge Transfer and Integration	.17*	.02
Board Engagement	.08	<.01
Internal Evaluator Engagement	.13*	.02
Internal Staff Engagement	.08	<.01

Note. * $p \leq .05$

** $p \leq .001$

¹³ The uniqueness index for a given variable is the percentage of variance accounted for by that variable, beyond the variance accounted for by the other predictor variables.

Table 14 shows that only Leadership Evaluation Characteristics, Knowledge Transfer and Integration, and Internal Evaluator Engagement displayed significant beta weights. Leadership Evaluation Characteristics displayed the largest beta weight at .23 ($p < .001$), while the beta weight for Knowledge Transfer and Integration was .17 ($p < .05$) and for Internal Evaluator Engagement was .13 ($p < .05$). Despite the significance of the beta weights, these variables individually accounted for very little unique variance (<5% of 35%) in Internal Learning, beyond the variance accounted for by the other variables.

For Accountability, the equation containing these significant six variables accounted for 23% of the variance, ($F(6, 251) = 12.49; p < .001$). Beta weights (standardized multiple regression coefficients) and uniqueness indices were then reviewed to assess the relative importance of the variables in the prediction of use of evaluation for Accountability. Beta weights (β) and uniqueness indices (r) are presented in Table 15.

Table 15

Factors Impacting Accountability

Predictor	Accountability	
	β	r
Leadership Evaluation Characteristics	.32**	.08
Managerial Commitment	-.01	<.01
Systems Perspective	-.04	<.01
Openness and Experimentation	.01	<.01
Knowledge Transfer and Integration	.09*	.03
Board Engagement	.00	<.01

Note. * $p \leq .05$

** $p \leq .001$

Table 15 shows that only Leadership Evaluation Characteristics and Knowledge Transfer and Integration displayed significant beta weights. Leadership Evaluation Characteristics displayed the largest beta weight at .32 ($p < .001$), uniquely accounting for 8% from the total variance of 29%, while the beta weight for Knowledge Transfer and Integration was .09 ($p < .05$), accounting for 3% of the total variance of 29%.

For Image Building, the equation containing these significant six variables accounted for 23% of the variance, ($F(6, 251) = 12.65; p < .001$). Beta weights (standardized multiple regression coefficients) and uniqueness indices were then reviewed to assess the relative

importance of the variables in the prediction of use of evaluation for Image Building. Beta weights (β) and uniqueness indices (r) are presented in Table 16.

Table 16

Factors Impacting Use for Image Building

Predictor	Image Building	
	β	r
Leadership Evaluation Characteristics	.27**	.06
Internal Evaluator	.17*	.03
Managerial Commitment	-.06	<.01
Systems Perspective	.04	<.01
Openness and Experimentation	.05	<.01
Knowledge Transfer and Integration	.23*	.03

Note. * $p \leq .05$

** $p \leq .001$

Table 16 shows that only Leadership Evaluation Characteristics, Internal Evaluator, and Knowledge Transfer and Integration displayed significant beta weights. Leadership Evaluation Characteristics displayed the largest beta weight at .27 ($p < .001$), while the beta weight for Knowledge Transfer and Integration was .23 ($p < .05$) and for Internal Evaluator was .17 ($p < .05$). Of the total variance of 23%, Leadership Evaluation Characteristics

uniquely accounted for 6%, Internal Evaluator for 3%, and Knowledge Transfer and Integration for 3%.

Across the three dependent variables of Use of Evaluation, three independent variables repeatedly emerge as playing the largest role in uniquely predicting Use of Evaluation: Leadership Evaluation Characteristics, Internal Evaluator Engagement, and Knowledge Transfer and Integration. Of these three, Leadership Evaluation Characteristics uniquely accounts for the most variance in all three dependent variables. These results must be interpreted cautiously, however, as there is some evidence of multicollinearity among the variables (Appendix G), compounded by the number of predictors included, which potentially biases regression coefficients by masking statistical significance. This reinforces the importance of using path analysis in the overall model as path analysis allows for covariation between all predicting variables and provides information about the relationships between multiple predictors and the dependent variables simultaneously (Keith, 2006). Consequently, while only three predictors retained significance in this combined model and will be used in the mediational model (research question 7), all predictors that were significantly correlated with the three dependent measures of Use of Evaluation in previous research questions will be included in the overall path analysis model.

Research Question 6: Is there a Correlation between Implementation of Evaluation and Use of Evaluation by the Nonprofit?

Pearson correlations were used to assess the direct relationships between the two measures of Implementation: Evaluation Activities and Evaluation Prevalence and the three

dependent variables of use of evaluation. As can be seen from Table 17, both Evaluation Activities and Evaluation Prevalence were significantly and positively correlated with Internal Learning, Accountability, and Image Building.

Table 17

Correlations between the Implementation Variables and Evaluation Use Measures

Variable	<i>Pearson Correlations</i>		
	Internal Learning	Accountability	Image Building
<i>Implementation</i>			
Evaluation Activities	.46**	.39**	.40**
Evaluation Prevalence	.50**	.40**	.37**

Note. * $p \leq .05$

** $p \leq .001$

Using multiple regression analyses, Internal Learning, Accountability, and Image Building scores were regressed on the linear combination of Evaluation Activities and Evaluation Prevalence. The equation containing these two variables accounted for 31% of the variance in Internal Learning, ($F(2, 255) = 58.38; p < .001$), 21% of the variance in Accountability, ($F(2, 255) = 33.35; p < .001$), and 20% of the variance in Image Building, ($F(2, 255) = 31.68; p < .001$). Beta weights (β) and uniqueness indices (r) are presented in Table 18. Table 18 shows that while correlated, both Evaluation Activities and Evaluation Prevalence displayed significant beta weights, uniquely accounting for up to 10% of the overall variance in each of the three dependent variables. Thus it can be concluded that

Evaluation Activities and Evaluation Prevalence are positively correlated with the three measures of Use of Evaluation and individually account for a significant portion of unique variance, and consequently, both variables will be retained for the multivariate analysis.

Table 18

Implementation Factors Impacting Evaluation Use

Variable	Internal Learning		Accountability		Image Building	
<i>Implementation</i>	β	r	β	r	β	r
Evaluation Activities	.28**	.10	.25**	.07	.29**	.04
Evaluation Prevalence	.37**	.06	.28**	.05	.23**	.07

Note. * $p \leq .05$

** $p \leq .001$

Research Question 7: Is the Combined Effect of Stakeholder Engagement, Leadership Evaluation Characteristics and the Dimensions Underlying Organizational Learning on the Dependent Variable of Use of Evaluation by the Nonprofit Mediated by Implementation?

As the two measures of Implementation, Evaluation Activities and Evaluation Prevalence, were positively and significantly correlated with the three measures of Use of Evaluation, it is important to assess whether these variables have an indirect effect on (mediate) the relationship between the significant predictors and the three dependent variables of Use of Evaluation. According to Barron and Kenny (1986) and Frazier, Tix, and Barron (2004) mediation can be said to occur when the independent variable significantly affects the mediator, the independent variable significantly affects the dependent variable in

the absence of the mediator, the mediator has a significant unique effect on the dependent variable, and the statistical significance of the independent variable on the dependent variable is reduced upon addition of the mediator to the model. As the analysis in research question six demonstrated that the two measures of Implementation had a significant and unique effect on the dependent variables of Use of Evaluation, what remained to be tested was the relationship between the independent variables and the mediator, and whether the relationship between independent and dependent variables is reduced by the addition of the two measures of Implementation to the model.

Using multiple regression analyses, Evaluation Activities and Evaluation Prevalence scores were regressed on the linear combination of Leadership Evaluation Characteristics, Board Engagement, Internal Evaluator Engagement, Internal Staff Engagement, and the four measures of Organizational Learning Capability. The equation containing these eight variables accounted for 30% of the variance in Evaluation Activities, ($F(8, 249) = 13.53$; $p < .001$) and 22% of the variance in Evaluation Prevalence, ($F(8, 249) = 8.62$; $p < .001$). Beta weights (β) are presented in Table 19 and show that only Leadership Evaluation Characteristics, Internal Staff Engagement, and Knowledge Transfer and Integration are potentially mediated by Evaluation Activities, while Evaluation Prevalence potentially mediates only Leadership Evaluation Characteristics.

Table 19

Factors Impacting Evaluation Activities and Evaluation Prevalence

	Evaluation Activities	Evaluation Prevalence
Predictor	β	β
Leadership Evaluation Characteristics	.41**	.32**
Managerial Commitment	.08	.14
Systems Perspective	-.07	.03
Openness and Experimentation	-.00	-.01
Knowledge Transfer and Integration	.15*	.02
Board Engagement	.00	.03
Internal Evaluator Engagement	.10	.10
Internal Staff Engagement	.13*	.07

Note. * $p \leq .05$

** $p \leq .001$

To assess whether the effect of Leadership Evaluation Characteristics, Internal Staff Engagement, and Knowledge Transfer and Integration on the three dependent variables shrinks upon addition of the two mediators to the model, I used the Aroian version of the Sobel test (Sobel, 1982; Aroian, 1944),¹⁴ suggested in Baron and Kenny (1986). This revealed that Evaluation Activities and Evaluation Prevalence mediated the relationships

¹⁴ $z\text{-value} = a*b/\text{SQRT}(b^2*s_a^2 + a^2*s_b^2 + s_a^2*s_b^2)$

between Leadership Evaluation Characteristics and Internal Learning, Accountability, and Image Building (Table 20). Consequently, while it was previously predicted that Implementation would mediate the relationships between all of the independent variables and the dependent variables of Use of Evaluation (Appendix A), the overall model will only position the measures of Implementation as mediating the relationships between Leadership Evaluation Characteristics and the three dependent variables of Use of Evaluation.

Table 20

Mediating Influence of Evaluation Activities (EA) and Evaluation Prevalence (EP)

Variable	<i>Z-scores</i>					
	Internal Learning		Accountability		Image Building	
	EA	EP	EA	EP	EA	EP
Leadership Evaluation Characteristics	4.04**	1.92*	2.98**	2.92**	2.31*	3.09**
Knowledge Transfer and Integration	0.18	0.28	1.55	0.32	1.61	0.31
Internal Staff Engagement	0.81	0.61	1.64	1.06	1.71	1.04

Note. * $p \leq .05$

** $p \leq .001$

Research Question 8: Are Any of the Correlations (Between Stakeholder Engagement, Leadership Evaluation Characteristics, the Dimensions Underlying Organizational Learning and Implementation) with Use of Evaluation by the Nonprofit Moderated by Organizational Characteristics, such as Size, Budget, Age, Funder, and/or Type of Organization?

To assess the potential moderating impact of Organizational Characteristics, univariate interactions were used for each dependent variable for each significant correlation. As Table 21 shows, only two Organizational Characteristics, Budget and Time Doing Evaluation, emerged as significant moderators. Budget moderated the relationship between Systems Perspective and Image Building, Openness and Experimentation, and both Internal Learning and Accountability, and Evaluation Activities and Internal Learning. Time Doing Evaluation moderated the relationships between Leadership Evaluation Characteristics and Image Building, Openness and Experimentation and Image Building, Knowledge Transfer and Integration and all three dependent variables, and Evaluation Activities and Internal Learning (Table 21).

Table 21

Moderators of Relationships between Predictor Variables and Evaluation Use Measures

Interactions	<i>F-Value</i>		
	Internal Learning	Accountability	Image Building
Leadership Evaluation Characteristics*Time Doing Evaluation			1.65*
Systems Perspective*Budget			2.04*
Openness and Experimentation*Time Doing Evaluation			1.39*
Openness and Experimentation*Budget	2.56*	1.89*	
Knowledge Transfer and Integration* Time Doing Evaluation	1.42*	1.59*	1.39*
Evaluation Activities*Time Doing Evaluation	1.43*		
Evaluation Activities*Budget	2.15*		

Note. * $p \leq .05$ ** $p \leq .001$

To view these interactions more clearly, these moderators were divided into groups based on a relatively even distribution. Budget was divided into three groups: less than \$250,000; \$250,000 to \$1,000,000 and over \$1,000,000; and Time Doing Evaluation was split into two groups: less than seven years and more than seven years. These Budget and Time Doing Evaluation groups were then used to graph each interaction (Appendix H). As Appendix H shows, while the moderating impact of Budget and Time Doing Evaluation was significant at $p < .05$, the influence of both of these characteristics on relationships between the independent and dependent variables was minimal. Consequently, these results should be interpreted cautiously, as while they showed significance, they were not powerful and could

be better interpreted as highlighting the structural characteristics that should be paid attention to in the subsequent analyses.

For the dependent variable of Internal Learning, interactions were graphed for: the effect of Evaluation Activities, conditional on Budget (Figure 27) and Time Doing Evaluation (Figure 37); the effect of Openness and Experimentation, conditional on Budget (Figure 29); the effect of Leadership Evaluation Characteristics, conditional on Budget (Figure 31); and the effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation (Figure 34). For larger nonprofits (over \$1,000,000 in budget), there was a stronger relationship between Evaluation Activities and Use of Evaluation for Internal Learning than for smaller (less than \$250,000) or medium (less than \$1,000,000) nonprofits. Compared to smaller nonprofits, the more the larger nonprofit engaged in Evaluation Activities, the more frequently they used the results for Internal Learning (Figure 27). This relationship was echoed in regards to Time Doing Evaluation; for nonprofits that had been doing evaluation for more than seven years, there was a stronger relationship between Evaluation Activities and Use of Evaluation for Internal Learning. Compared to nonprofits doing evaluation for less than seven years, the more nonprofits that had been doing evaluation for more than seven years engaged in Evaluation Activities, the more they used results for Internal Learning (Figure 37). While the remaining Internal Learning interactions: the effect of Openness and Experimentation, conditional on Budget (Figure 29); the effect of Leadership Evaluation Characteristics, conditional on Budget (Figure 31); and the effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation (Figure 34) were

significant, these graphs show only minimal differences between groups and should be interpreted cautiously.

For the dependent variable of Accountability, interactions were graphed for the effect of Openness and Experimentation, conditional on Budget (Figure 28) and the effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation (Figure 36). While small, medium, and large nonprofits differed slightly in regards to the effect of Openness and Experimentation on use of evaluation for Accountability, the general trend was that that the higher these nonprofits scored on the dimension of Openness and Experimentation, the more they used evaluation for Accountability, with smaller nonprofits displaying a slightly stronger effect (Figure 28). For nonprofits that had been doing evaluation for less than seven years, there was a stronger positive effect of Knowledge Transfer and Integration on use of evaluation for Accountability. The higher these nonprofits scored on the dimension of Knowledge Transfer and Integration, the more they used evaluation for Accountability (Figure 36).

For the dependent variable of Image Building, interactions were graphed for: the effect of Systems Perspective, conditional on Budget (Figure 30); the effect of Leadership Evaluation Characteristics, conditional on Time Doing Evaluation (Figure 32); the effect of Openness and Experimentation, conditional on Time Doing Evaluation (Figure 33); and the effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation (Figure 36). For smaller and medium nonprofits, there was a steeper positive increase between scores on the dimension of Systems Perspective and use of evaluation for Image Building. The higher smaller and medium nonprofits scored on Systems Perspective, the more they

used evaluation for Image Building (Figure 30). The interaction between Leadership Evaluation Characteristics and Time Doing Evaluation revealed a stronger impact of Leadership Evaluation Characteristics on use of evaluation for Image Building for nonprofits that had been conducting evaluation for more than seven years (Figure 32). In contrast, nonprofits that had been conducting evaluation for less than seven years showed a steeper positive increase between scores on the dimension of Openness and Experimentation and use of evaluation for Image Building (Figure 33). While significant, the graphs of the interaction between Knowledge Transfer and Integration and Time Doing Evaluation showed minimal differences between groups, with those conducting evaluation for less time scoring slightly lower but still maintaining the same positive relationship (Figure 36).

Overall Model

The overall path analysis model was constructed based on the results from the research questions (Figure 17). Data were analyzed using the SAS System's CALIS procedure (SAS Institute Inc., 1989), and the models tested were covariance structure models with multiple indicators for all latent constructs. Standard deviations and intercorrelations for the variables are presented in Appendices E, F, and G.

The present analysis followed a two-step procedure based in part on an approach recommended by Anderson and Gerbing (1988). In the first step, confirmatory factor analysis was used to develop a measurement model that demonstrated an acceptable fit to the data. In step two, the measurement model was modified so that it came to represent the theoretical (causal) model of interest. This theoretical model was then tested to assess whether a theoretically meaningful and statistically acceptable model was found.

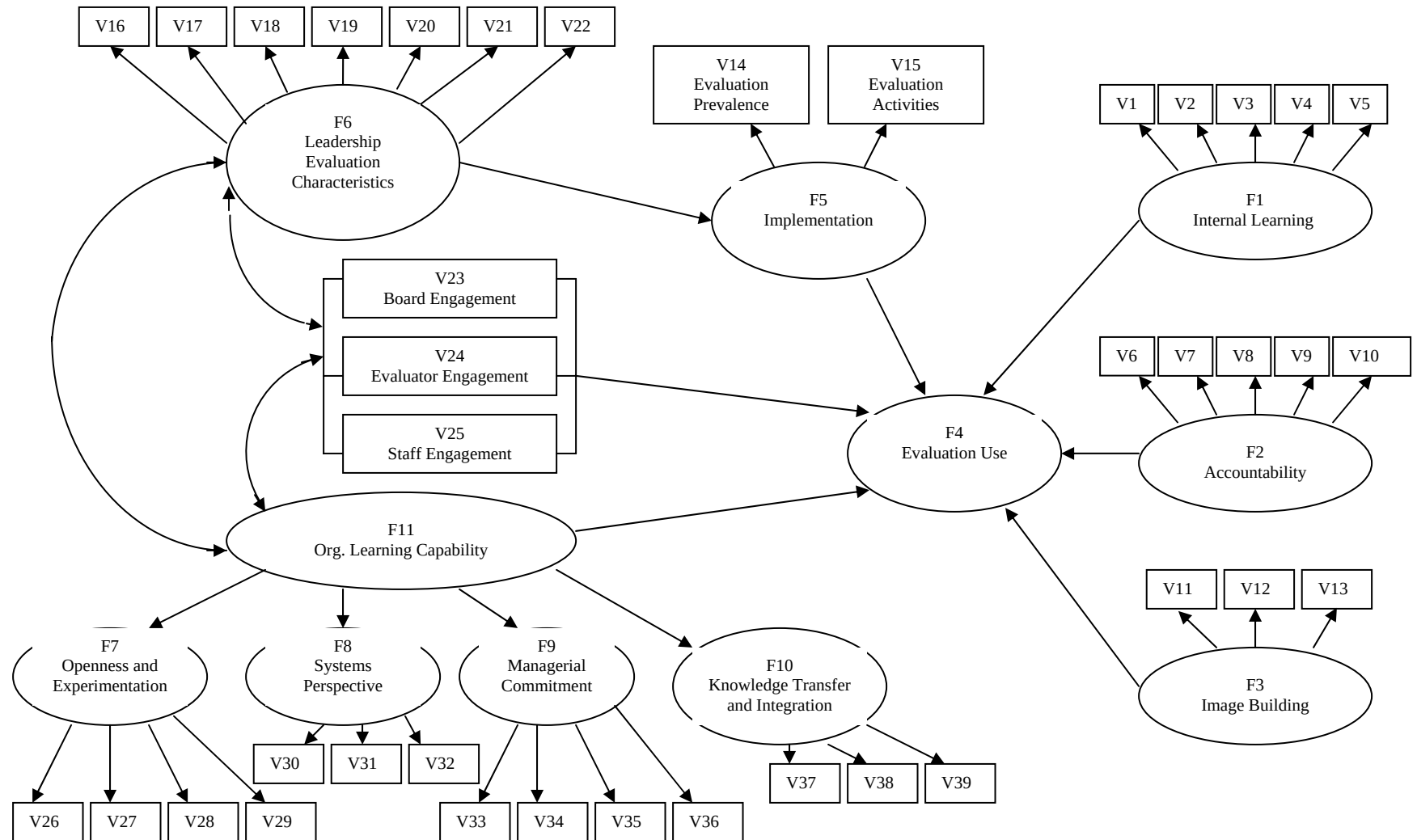


Figure 17: Overall path analysis model.

The Measurement Model

In path analysis with latent variables, a measurement model describes the nature of the relationship between (a) a number of latent variables, or factors, and (b) the manifest indicator variables that measure those latent variables. The model investigated in this research consisted of thirteen latent and manifest variables corresponding to the thirteen constructs of the Use of Evaluation for Organizational Learning model: Leadership Evaluation Characteristics; three measures of Stakeholder Engagement: Board Engagement, Internal Evaluator Engagement, Internal Staff Engagement; four measures of Organizational Learning Capability: Managerial Commitment, Systems Perspective, Openness and Experimentation, and Knowledge Transfer and Integration; two measures of Implementation: Evaluation Prevalence and Evaluation Activities; and three measures of Evaluation Use: Internal Learning, Accountability, and Image Building. Each of these latent variables was measured by at least three manifest indicator variables (Figure 17).

The Initial Measurement Model

Figure 17 follows Bentler's (1989) convention of identifying latent variables or "factors" with the letter "F" and labeling manifest variables with the letter "V." Eleven latent constructs, as well as the indicators that measure these constructs, in addition to five manifest variables were investigated in this study. Figure 17 shows that Evaluation Use (F4) is measured by the three dependent variables of Use of Evaluation (F1, F2, and F3), which are measured by manifest variables V1 to V13. Leadership Evaluation Characteristics (F6) is measured by V16 to V22, Implementation (F5) by Evaluation Prevalence (V14) and

Evaluation Activities (V15) and Organizational Learning Capability (F11), by the four latent factors (F7, F8, F9, F10), which are measured by V26 to V39.

This measurement model was estimated using the maximum likelihood method, and the chi-square value was statistically significant, $X^2(48, N=258) = 109.37, p < .0001$. While the chi-square statistic may technically be used to test the fit of the model to the data, in practice, this statistic is very sensitive to sample size and will often result in the rejection of a well-fitting model (Jöreskog & Sörbom, 1989; Hatcher, 1994). Consequently, while the chi-square estimate was significant, the initial measurement model (Figure 17) was indicative of an acceptable fit (GFI=.93, RMSEA=.07, NFI=.92, NNFI=.91, CFI=.95).

Standardized factor loadings for the indicator variables are presented in Table 22. The SAS CALIS procedure provides approximate standard errors for these coefficients which allow *t*-tests of the null hypothesis that the coefficients are equal to zero in the nonprofit population. The *t* scores obtained for the coefficients in Table 22 range from 10.74 to 17.12, indicating that all factor loadings were significant ($p < .001$). In line with Anderson and Gerbing (1988), this finding provides evidence supporting convergent validity. These findings, in addition to the variance explained (Table 22) provide some structural support for the validity of the constructs and their indicators. This measurement model was therefore retained as the final measurement model against which the theoretical models would be compared.

Table 22

Properties of the Measurement Model

Construct and Indicators	Standardized Loading	<i>t</i> -test*	Variance Explained
Evaluation Use (F4)			.59
Internal Learning (F1)	.87	16.13	.75
Accountability (F2)	.76	13.59	.58
Image Building (F3)	.66	11.11	.43
Organizational Learning Capability (F11)			.61
Managerial Commitment (F9)	.84	16.13	.72
Systems Perspective (F8)	.73	13.07	.53
Openness and Experimentation (F7)	.88	17.12	.77
Knowledge Transfer and Integration (F10)	.63	10.74	.40
Implementation (F5)			.49
Evaluation Prevalence (V14)	.68	10.91	.46
Evaluation Activities (V15)	.72	11.62	.52

Note. *All *t*-tests were significant at $p \leq .001$

The Theoretical Model

The theoretical model tested was identical to the final measurement model with manifest and latent variables in Figure 17 and, while the chi-square value was statistically significant, $X^2(50, N=258) = 127.01, p < .0001$, the goodness of fit statistics were indicative of an acceptable fit (GFI=.93, RMSEA=.07, NFI=.90, NNFI=.90, CFI=.94). The distribution of normalized residuals was fairly symmetrical and centered around zero, and no residuals exceeded 2 in absolute magnitude.

A review of the model showed that three of the causal paths linking the three measures of Stakeholder Engagement with Use of Evaluation proved to be non-significant at $p < .05$ (Table 23). The standardized path coefficients for the paths from Board Engagement (V23), Evaluator Engagement (V24), and Staff Engagement (V25) to Evaluation Use (F4) were all less than .030, *ns*.

The nomological validity of a theoretical model can be tested by performing a chi-square difference test in which the theoretical model is compared to the measurement model. A finding of no significant difference indicates that the theoretical model is successful in accounting for the observed relationships between the latent constructs (Anderson & Gerbing, 1988). Subtracting the chi-square for the measurement model from the theoretical model results in a difference value of $127.01 - 109.37 = 17.64$. The degrees of freedom (*df*) for the test are equal to the *df* difference between the two models, $50 - 48 = 2$. The critical chi-square value with 2 *df* is 13.82 ($p < .001$). This finding shows that the theoretical model was unsuccessful in accounting for the relationships between the latent constructs.

The non-significant causal paths in addition to the significant chi-square test of nomological validity showed that this initial model did not provide an acceptable fit to the data. Therefore, a specification search was conducted to ascertain a better fitting model.

Revised Model

MacCallum, Roznowski, and Necowitz (1992) caution against conducting a specification search based on a relatively small sample, as there is a possibility that data-driven modifications will capitalize on chance characteristics of the sample data and result in a final model that will not generalize to other nonprofits. Consequently, I started the search

by reviewing both Wald tests and Lagrange multiplier tests to identify parameters that could be modified and supported theoretically (Bentler & Bonett, 1980; Bentler & Chou, 1987).

A Wald test (Bentler & Bonett, 1980) suggested that it would be beneficial to delete the paths from Board Engagement (V23), Evaluator Engagement (V24), and Staff Engagement (V25) to Evaluation Use (F4), while a Lagrange multiplier test estimated a causal path should be added between these three measures of Stakeholder Engagement and the mediating factor of Implementation (F5). The addition of these paths would be consistent with the initial proposed model of the factors associated with the implementation and use of evaluation for organizational learning as outlined in Appendix A, suggesting that the more stakeholders were engaged in evaluation, the more evaluation would be implemented and therefore used.

When this resulting revised model was estimated, fit indices showed that while the chi-square value remained statistically significant, $X^2(50, N=258) = 119.12, p < .0001$, it had been reduced by the addition of these new paths. The goodness of fit statistics also slightly improved (i.e., the NFI and NNFI) and were indicative of an acceptable fit (GFI=.93, RMSEA=.07, NFI=.91, NNFI=.91, CFI=.94). To assess whether the revised model accounted for the relationships in the latent constructs, the chi-square difference test was again calculated comparing the measurement to the revised model. Subtracting the chi-square for the measurement model from the theoretical model resulted in a difference value of $119.12 - 109.37 = 9.75$, which, with 2 *df* was non-significant ($p > .001$). The non-significant chi-square indicated that the revised model provided a fit that was not significantly worse than that provided by a measurement model in which all variables were free to covary. This

finding supported that the revised model was successful in accounting for the observed relationship between the latent constructs.

Table 23 and Figure 18 display standardized path coefficients for the revised model. It can be seen that Organizational Learning Capability, Leadership Evaluation Characteristics, Staff Engagement, Evaluator Engagement, and Implementation coefficients were significant and in the predicted direction, while Board Engagement was not significant. R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 62 percent of the variance in Evaluation Use, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 47 percent of the variance in Implementation.

Table 23

Properties of the Theoretical and Revised Model

Dependent Variable/Independent Variable	Theoretical Model	Revised Model
Evaluation Use (F4)		
Organizational Learning Capability (F11)	.30*	.30*
Implementation (F5)	.63*	.64*
Board Engagement (V23)	.04	-
Evaluator Engagement (V24)	.09	-
Staff Engagement (V25)	-.03	-
Implementation (F5)		
Leadership Evaluation Characteristics (F6)	.75*	.61*
Board Engagement (V23)	-	.06
Evaluator Engagement (V24)	-	.18*
Staff Engagement (V25)	-	.13*

Note. * $p \leq .05$

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Comparisons between the Three Measures of Evaluation Use

To look more closely at the differential relationships between the constructs of the Use of Evaluation for Organizational Learning model and the three measures of Evaluation Use, this path analysis model was run for each measure of Evaluation Use: Internal Learning, Accountability, and Image Building. Figures 19, 20, and 21, and Table 24 display the standardized path coefficients for the revised model for each measure of Evaluation Use.

Internal Learning

When the revised path analysis model was estimated for the use of evaluation for Internal Learning, fit indices showed that while the chi-square value remained statistically significant, $X^2 (31, N=258) = 74.19, p < .0001$, the goodness of fit statistics were indicative of an acceptable fit (GFI=.95, RMSEA=.07, NFI=.91, NNFI= .92, CFI=.95). Figure 19 displays the standardized path coefficients for the revised Internal Learning model. It can be seen that Organizational Learning Capability, Leadership Evaluation Characteristics, Staff Engagement, Evaluator Engagement, and Implementation coefficients were significant and in the predicted direction, while Board Engagement was not significant. R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 49 percent of the variance in use of evaluation for Internal Learning, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 47 percent of the variance in Implementation.

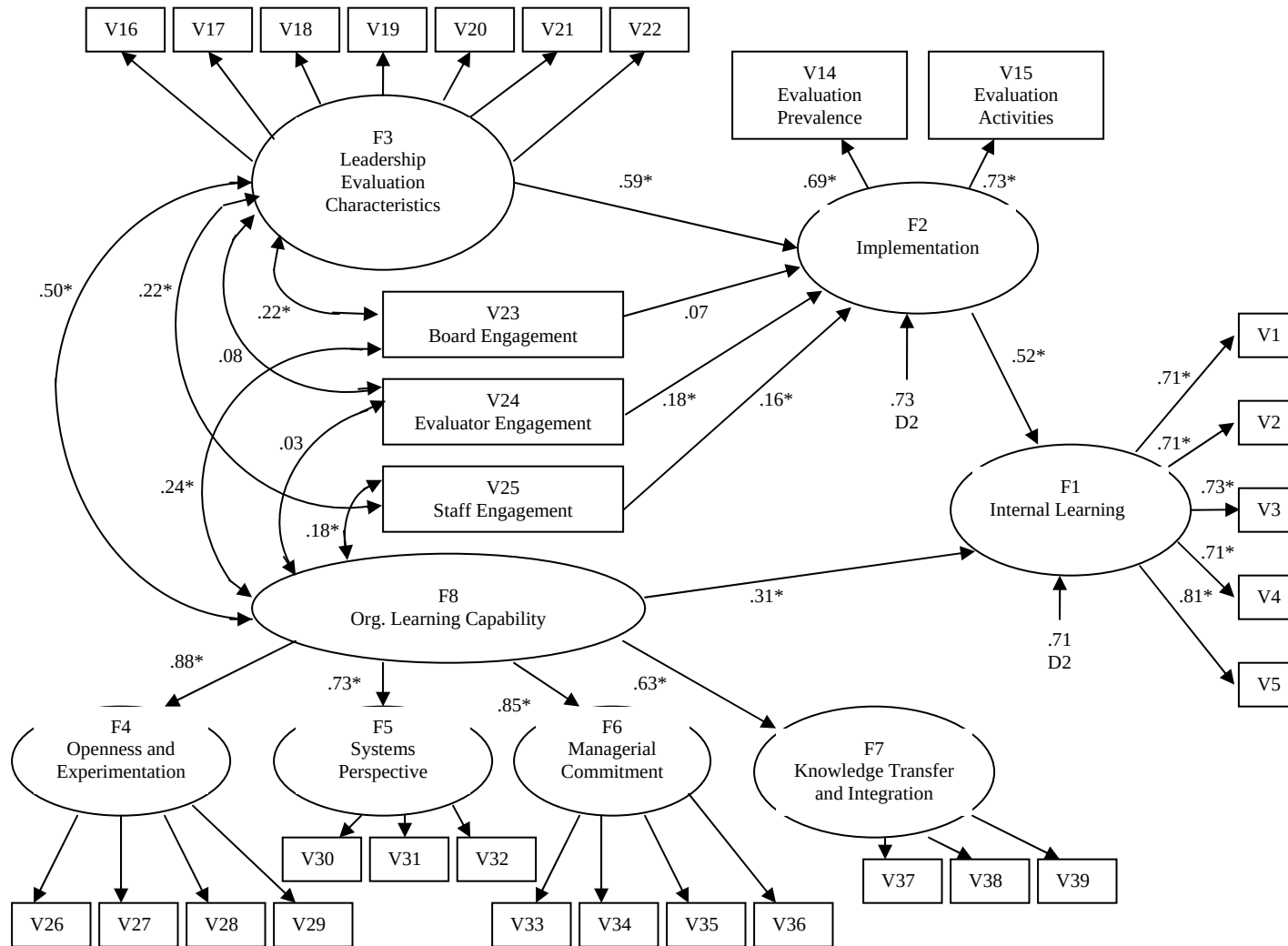


Figure 19: Revised path analysis model for Internal Learning.

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Accountability

When the revised path analysis model was estimated for the use of evaluation for Accountability, fit indices showed that while the chi-square value remained statistically significant, $X^2 (31, N=258) = 76.49, p<.0001$, the goodness of fit statistics were indicative of an acceptable fit (GFI=.95, RMSEA=.07, NFI=.90, NNFI= .91, CFI=.95). Figure 20 displays the standardized path coefficients for the revised Accountability model. It can be seen that Organizational Learning Capability, Leadership Evaluation Characteristics, Staff Engagement, Evaluator Engagement, and Implementation coefficients were significant and in the predicted direction, while Board Engagement was not significant. R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 32 percent of the variance in use of evaluation for Accountability, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 50 percent of the variance in Implementation

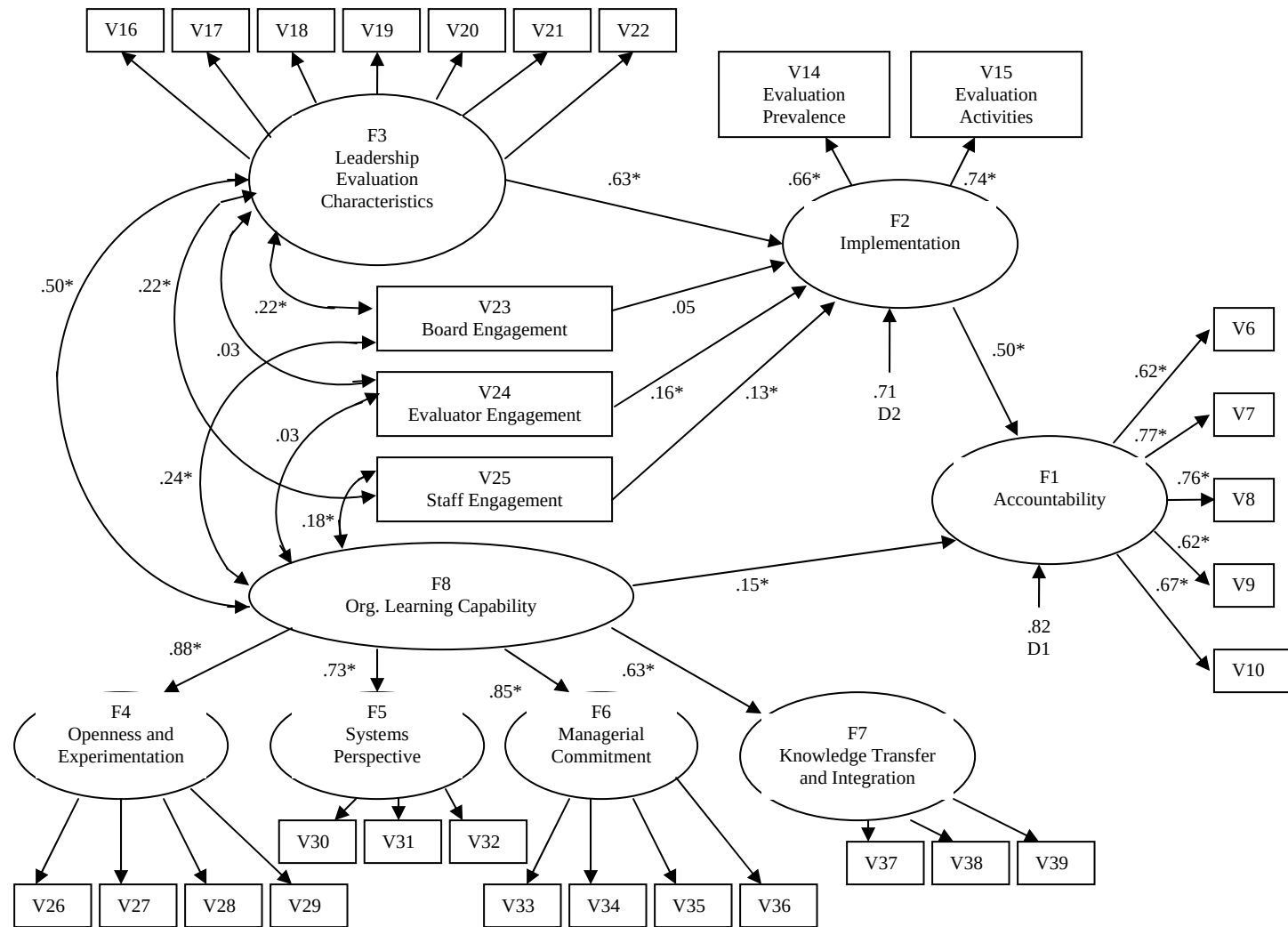


Figure 20: Revised path analysis model for Accountability.

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Image Building

When the revised path analysis model was estimated for the use of evaluation for Image Building, fit indices showed that while the chi-square value remained statistically significant, $X^2(31, N=258) = 86.79, p < .0001$, the goodness of fit statistics were indicative of an acceptable fit (GFI=.94, RMSEA=.08, NFI=.90, NNFI=.90, CFI=.93). Figure 21 displays the standardized path coefficients for the revised Image Building model. It can be seen that Organizational Learning Capability, Leadership Evaluation Characteristics, Staff Engagement, Evaluator Engagement, and Implementation coefficients were significant and in the predicted direction, while Board Engagement was not significant. R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 30 percent of the variance in use of evaluation for Image Building, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 49 percent of the variance in Implementation

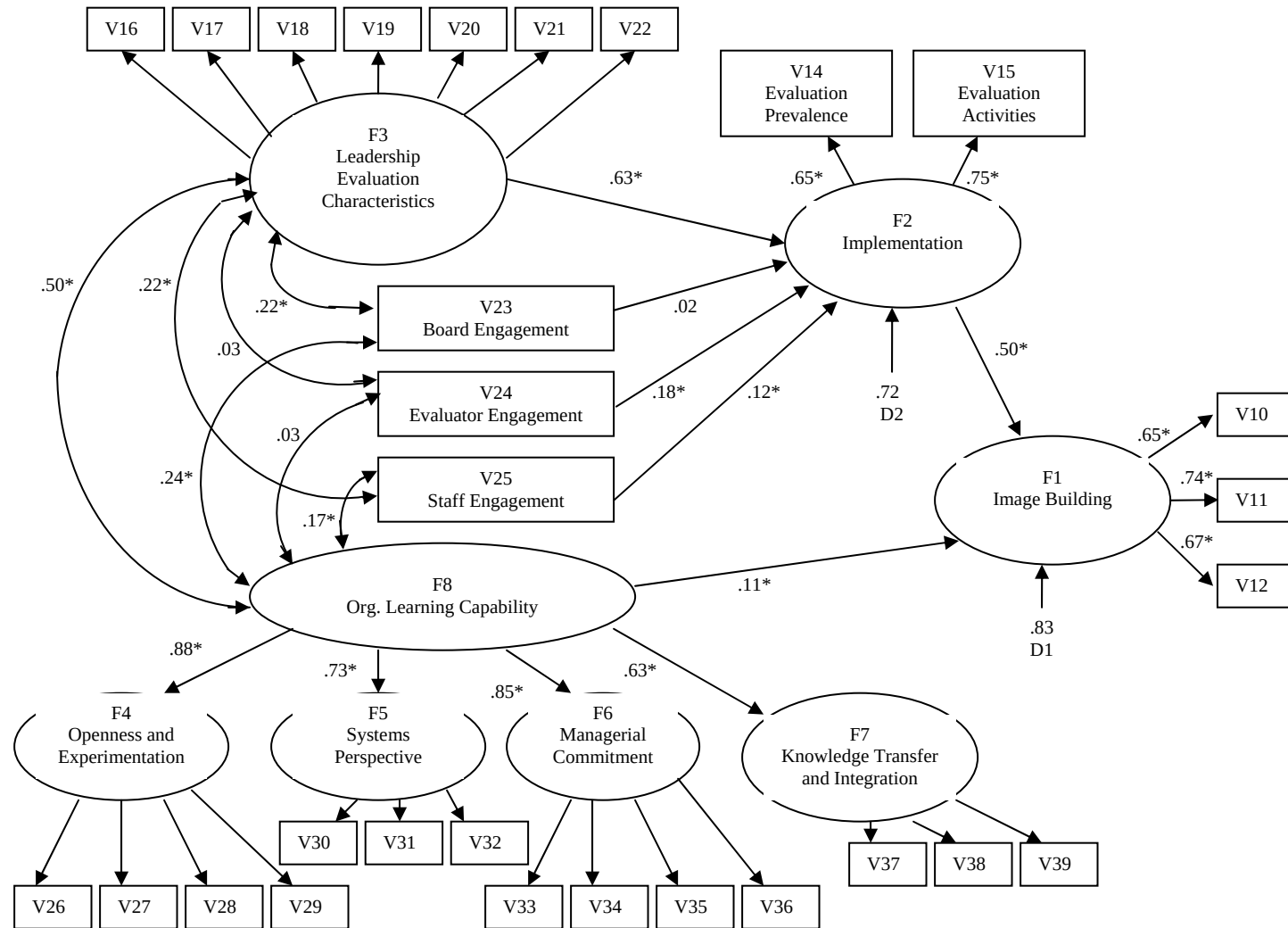


Figure 21: Revised path analysis model for Image Building.

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Differences between Evaluation Use Measures

As Table 24 clearly shows, while the overall revised path analysis model provides an acceptable fit for each measure of Evaluation Use, there are slight differences between the effects of variables on each measure of Evaluation Use. The most noteworthy difference is that while significant for all three measures of Evaluation Use, Organizational Learning Capability has a significantly stronger impact on whether a nonprofit uses evaluation for Internal Learning, than for Accountability or Image Building. Leadership Evaluation Characteristics remains the most significant predictor of Implementation, while Implementation is strongly correlated with Evaluation Use and acts as a mediator between Evaluation Use and Leadership Evaluation Characteristics, Evaluator Engagement, and Staff Engagement.

Table 24

Properties of the Revised Model for Each Measure of Evaluation Use

Dependent Variable/Independent Variable	Internal Learning	Accountability	Image Building
Evaluation Use			
Organizational Learning Capability	.31*	.15*	.11*
Implementation	.52*	.50*	.50*
Board Engagement	-	-	-
Evaluator Engagement	-	-	-
Staff Engagement	-	-	-
Implementation			
Leadership Evaluation Characteristics	.59*	.63*	.63*
Board Engagement	.07	.06	.02
Evaluator Engagement	.18*	.16*	.18*
Staff Engagement	.16*	.13*	.12*

Note. * $p \leq .05$

Moderating Impact of Budget

To look more closely at the differential relationships between the constructs of the Use of Evaluation for Organizational Learning model and the three measures of Evaluation Use when the moderating impact of Budget is included, this path analysis model was run for each level of Budget: less than \$250,000; \$250,000 to \$1,000,000 and over \$1,000,000. Figures 22, 23, and 24, and Table 25 display the standardized path coefficients for the path analysis model for each level of Budget.

Small Budget

For nonprofits with a budget of less than \$250,000, the only significant coefficient in the predicted direction was Organizational Learning Capability, while Board Engagement, Leadership Evaluation Characteristics, Staff Engagement, Evaluator Engagement, and Implementation coefficients were not significant (Figure 22). R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 26 percent of the variance in Evaluation Use, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 42 percent of the variance in Implementation (Table 25).

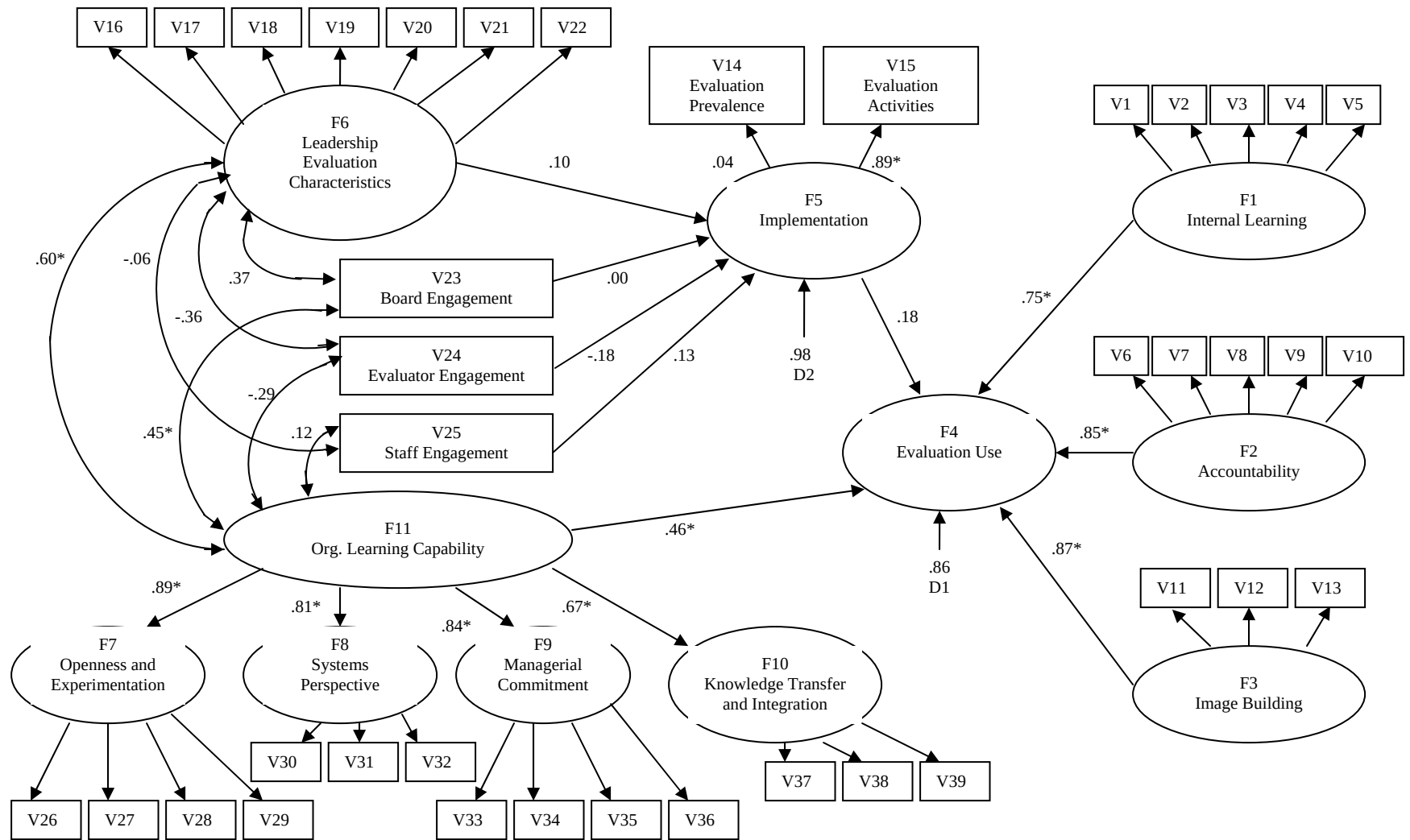


Figure 22: Moderating impact of small budget (<\$250,000).

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Medium Budget

For nonprofits with a budget of between \$250,000 and \$1,000,000, the Organizational Learning Capability, Implementation, and Leadership Evaluation Characteristics coefficients were significant and in the predicted direction, while Board Engagement, Staff Engagement, and Evaluator Engagement coefficients were not significant (Figure 23). R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 71 percent of the variance in Evaluation Use, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 30 percent of the variance in Implementation (Table 25).

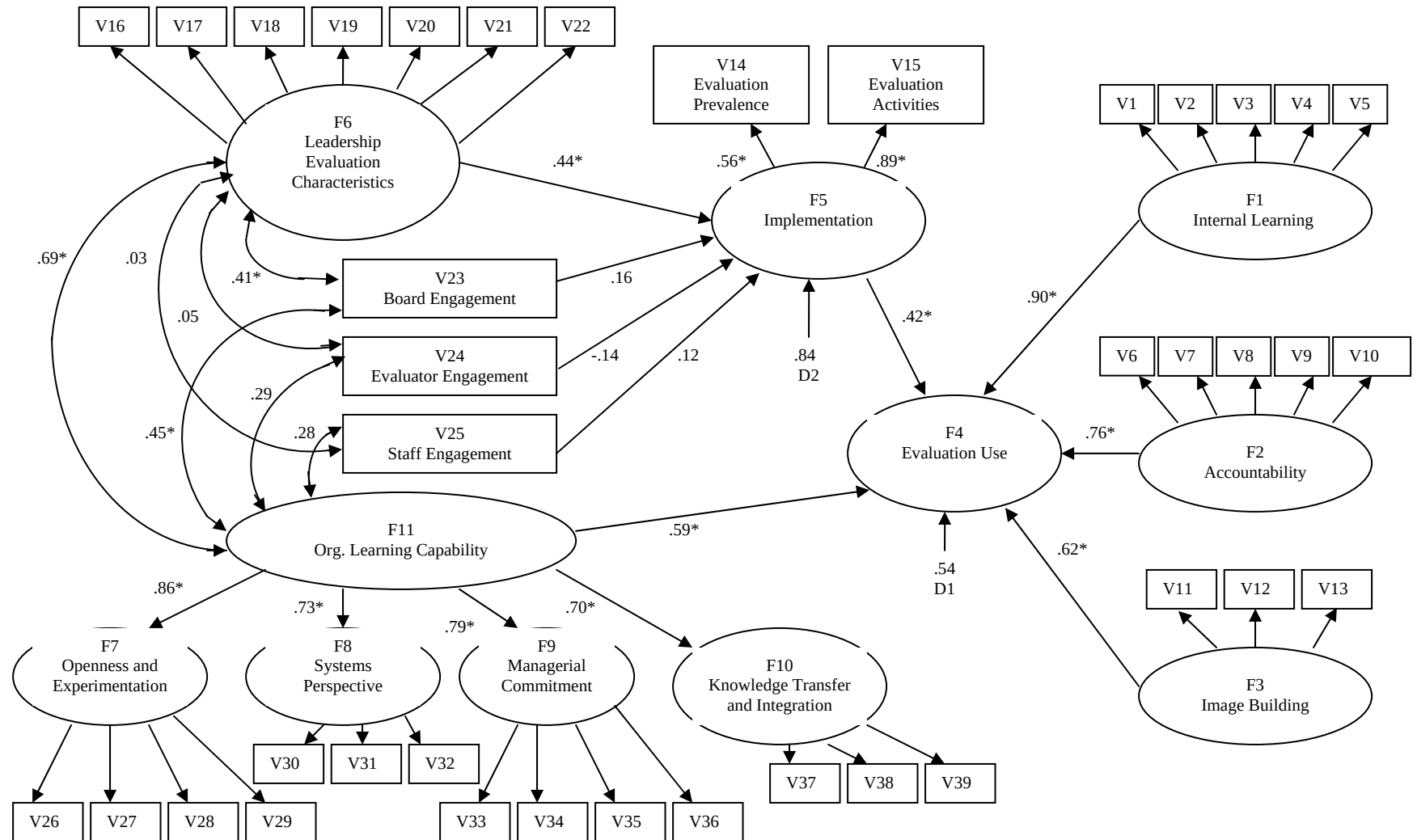


Figure 23: Moderating impact of medium budget (\$250,000-\$1,000,000).

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Large Budget

For nonprofits with a budget of more than \$1,000,000, the Implementation, and Leadership Evaluation Characteristics coefficients were significant and in the predicted direction, while Organizational Learning Capability, Board Engagement, Staff Engagement, and Evaluator Engagement coefficients were not significant (Figure 24). R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 48 percent of the variance in Evaluation Use, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 84 percent of the variance in Implementation (Table 25).

Table 25

Properties of the Moderating Impact of Budget Models

Dependent Variable/Independent Variable	Small Budget	Medium Budget	Large Budget
Evaluation Use (F4)	$r^2=.26$	$r^2=.71$	$r^2=.48$
Organizational Learning Capability (F11)	.46*	.59*	-.02
Implementation (F5)	.18	.42*	.70*
Implementation (F5)	$r^2=.42$	$r^2=.30$	$r^2=.84$
Leadership Evaluation Characteristics (F6)	.10	.44*	.80*
Board Engagement (V23)	.00	.16	.22
Evaluator Engagement (V24)	.18	-.14	.00
Staff Engagement (V25)	.13	.12	.09

Note. * $p \leq .05$

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Differences between Budget Groups

As Table 25 clearly shows, there is a moderating impact of Budget as there are group differences between the effects of variables on Evaluation Use. One noteworthy difference is that while Organizational Learning Capability is strongly associated with Evaluation Use for both small and medium budget nonprofits, it does not have a significant association for large budget nonprofits. In addition, while Leadership Evaluation Characteristics are positively associated with Implementation for both medium and large budget nonprofits, there is no significant relationship for small budget nonprofits. Implementation is not significantly associated with Evaluation Use for small budget nonprofits and thus does not act as a mediator.

Moderating Impact of Time Doing Evaluation

To look more closely at the differential relationships between the constructs of the Use of Evaluation for Organizational Learning model and the three measures of Evaluation Use when the moderating impact of Time Doing Evaluation is included, this path analysis model was run for two groups: nonprofits doing evaluation for less than seven years and nonprofits doing evaluation for more than seven years. Figures 25, 26, and Table 26 display the standardized path coefficients for the path analysis model for each group.

Less Time Doing Evaluation

For nonprofits doing evaluation for less than seven years, the only significant coefficients in the predicted direction were Organizational Learning Capability and Leadership Evaluation Characteristics, while Board Engagement, Staff Engagement, Evaluator Engagement, and Implementation coefficients were not significant (Figure 25). R^2

values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 67 percent of the variance in Evaluation Use, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 63 percent of the variance in Implementation (Table 26).

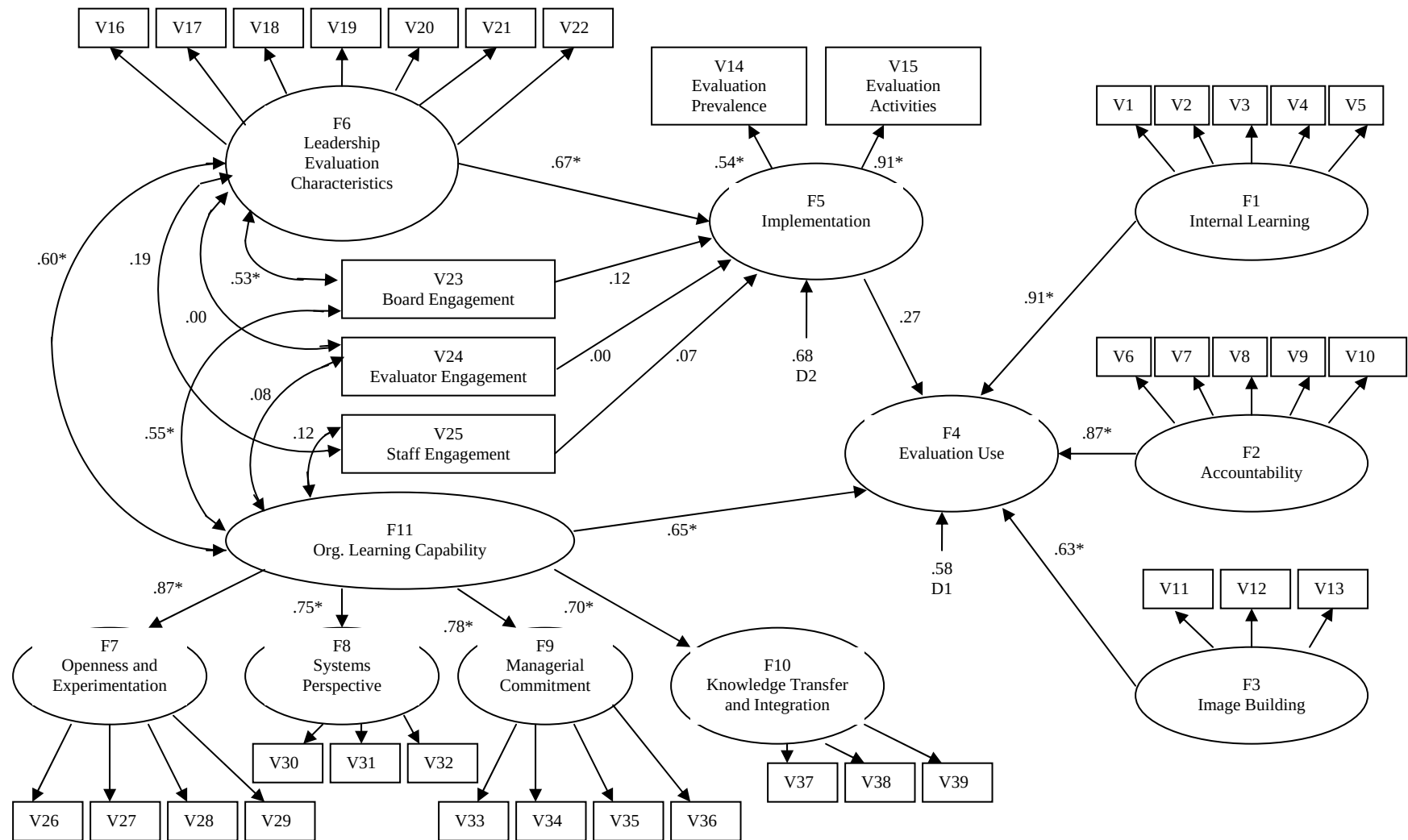


Figure 25: Moderating impact of less time doing evaluation.

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

More Time Doing Evaluation

For nonprofits doing evaluation for more than seven years, Implementation and Leadership Evaluation Characteristics coefficients were significant and in the predicted direction, while Organizational Learning Capability, Board Engagement, Staff Engagement, Evaluator Engagement, and coefficients were not significant (Figure 26). R^2 values showed that Organizational Learning Capability, Leadership Evaluation Characteristics, and Implementation accounted for 80 percent of the variance in Evaluation Use, while Leadership Evaluation Characteristics, Staff Engagement, and Evaluator Engagement accounted for 62 percent of the variance in Implementation (Table 26).

Table 26

Properties of the Moderating Impact of Time Doing Evaluation (TDE) Models

Dependent Variable/Independent Variable	Less TDE	More TDE
Evaluation Use (F4)	$r^2=.67$	$r^2=.80$
Organizational Learning Capability (F11)	.65*	.08
Implementation (F5)	.27	.86*
Implementation (F5)	$r^2=.53$	$r^2=.62$
Leadership Evaluation Characteristics (F6)	.67*	.72*
Board Engagement (V23)	.12	.05
Evaluator Engagement (V24)	.00	.02
Staff Engagement (V25)	.07	.26

Note. * $p \leq .05$

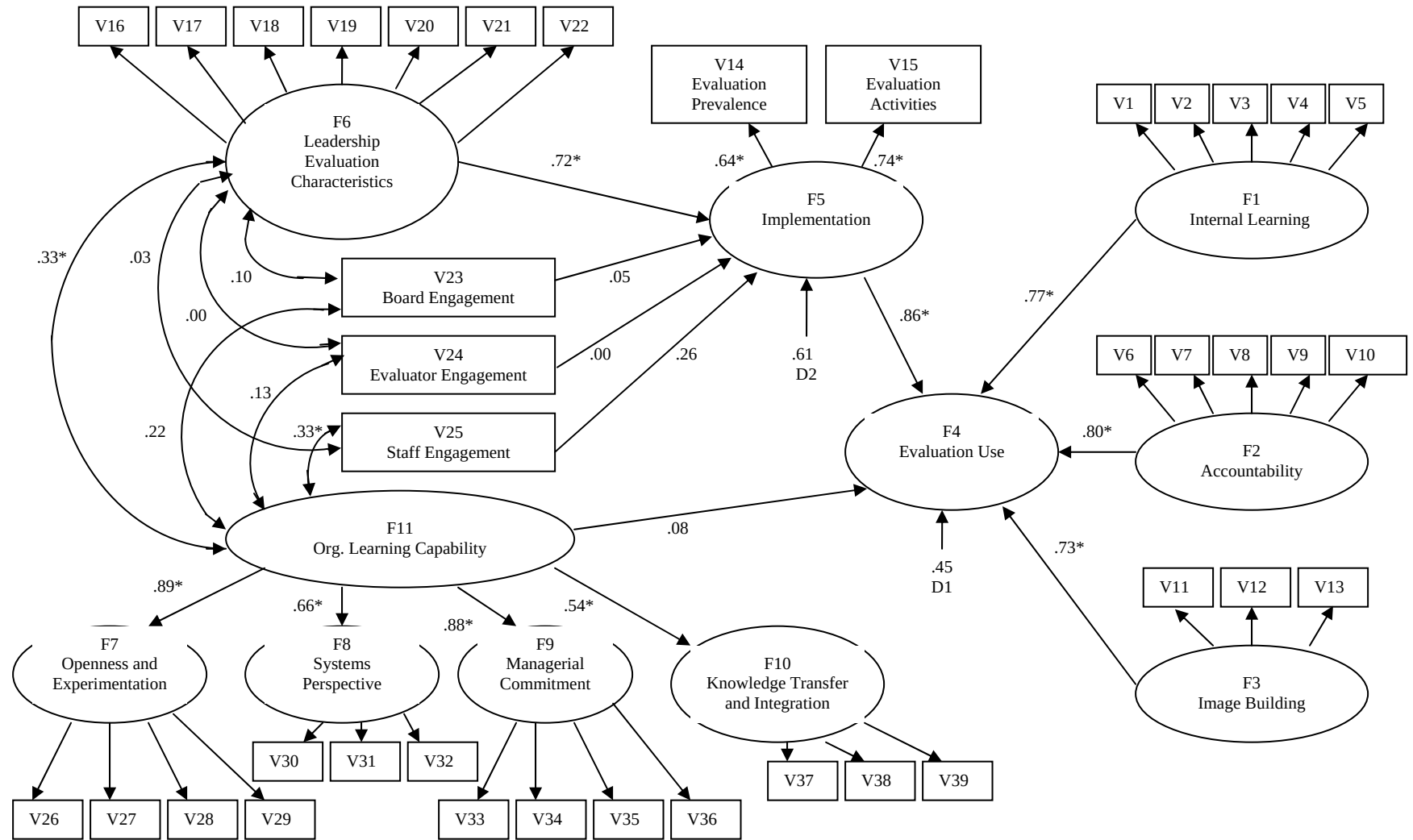


Figure 26: Moderating impact of more time doing evaluation.

Note: Standardized path coefficients appear on single-headed arrows, while correlations appear on curved, double-headed arrows.

* $p \leq .05$

Differences between Time Doing Evaluation Groups

As Table 26 clearly shows, there is a moderating impact of Time Doing Evaluation as there are group differences between the effects of variables on Evaluation Use. One noteworthy difference is that while Leadership Evaluation Characteristics are positively associated with Implementation for both groups, Implementation is not significantly associated with Evaluation Use for nonprofits doing evaluation for less than seven years and thus does not act as a mediator. In addition, for nonprofits doing evaluation for less than seven years, Organizational Learning Capability is significantly associated with Evaluation Use, while there is no significant relationship for nonprofits doing evaluation for more than seven years.

VI: DISCUSSION

If evaluation is to be more useful to nonprofits, the nonprofit sector needs to understand how evaluation can contribute to organizational learning and improve programs and practices, rather than be an external tool for accountability. A major deficiency in attempts to move toward this is the lack of empirical data on organizational factors associated with the implementation and use of evaluation for organizational learning. Drawing on the existing literature and using a large and diverse sample of nonprofits from across North Carolina, this dissertation set out to explore the interconnections between Evaluation Use, specifically, use for Internal Learning, Accountability, and Image Building and several organizational characteristics: Leadership Evaluation Characteristics, Stakeholder Engagement, and Organizational Learning Capability.

Review of Previous Findings

A review of the existing empirical research surrounding the factors associated with the implementation and use of evaluation for organizational learning revealed a general lack of solid empirical work, particularly in regard to the nonprofit sector. Evaluation researchers, such as Cousins et al. (2004), Leviton (2003), Torres and Preskill (2001), and Weiss (1988a, 1998) critiqued the paucity of research surrounding evaluation use, which is dominated by individual case study approaches and small sample sizes, reducing the power of any analysis and limiting any potential for replication. The nonprofit management literature recognizes that evaluation approaches in the for-profit sector may not be applicable to nonprofits, yet little data is available on nonprofit evaluation and effectiveness (Drucker, 1990; Forbes,

1998). While there are generic descriptions of nonprofit evaluation practice, these are primarily focused on case study approaches with descriptive and prescriptive results (Fine, Thayer, & Coghlan, 1998; McNelis & Bickel, 1996; Morley, Hatry, & Cowan, 2002; Morley, Vinson, & Hatry, 2001; Poole et al., 2000; Sawhill & Williamson, 2001; U.S. G.A.O, 1998).

In addition to the lack of empirical research surrounding evaluation practice, there is a lot of theory about the organizational and environmental factors that may impact the quality of evaluation in the nonprofit sector, but no empirical evidence base on the importance of these factors, on how they interact, or on how they impact nonprofits' implementation and use of evaluation to benefit programs and build nonprofit capacity (Hatry, van Houten, Plantz, & Greenway, 1997; Newcomer, 1997; Taylor & Sumariwalla, 1993). While there is some agreement that stakeholder engagement can improve evaluation processes in general (Brandon, 1998; Greene, 1988a, 1988b; Turnbull, 1999), as well as in the nonprofit sector (Fine, Thayer, & Coghlan, 1998, 2000), what is unknown is how correlated stakeholder engagement is with use of evaluation. While qualitative research emphasized the importance of supportive nonprofit leadership in the evaluation process (Alaimo & Van Slyke, 2006), no quantitative data has looked at the role of nonprofit leadership in relation to evaluation use. While research has been conducted looking at direct correlations of organizational learning and performance (Ellinger, Yang, & Ellinger, 2000; Zhang, Zhang, & Yang, 2004; Ismail, 2005), no research has looked at the importance of organizational learning capability in regards to use of evaluation by nonprofits.

This dissertation aimed to provide a useful foundation for future research into issues of using evaluation to improve programs and practices, and ultimately, to benefit our

communities. By using a large and diverse sample of nonprofits, this quantitative study attempted to redress some of the methodological imbalances, providing more robust measures of organizational constructs and use of evaluation, in addition to using more powerful statistical tools, such as multivariate analysis, to explore relationships between factors highlighted as impacting use of evaluation for organizational learning in the nonprofit sector. While there are clearly limitations to any empirical work, this study also provided a clearer model or framework in which to operationalize and assess the factors that impact nonprofits using evaluation for organizational learning.

Review of Main Findings

In examining the relationships between organizational characteristics and use of evaluation by the nonprofit, several issues emerged. How use of evaluation is operationalized and measured is critical to understanding what factors may facilitate a nonprofit's own use of evaluation and developing instrumental scales of use may contribute to a greater understanding of this process. Building on this, while Leadership Evaluation Characteristics, Evaluator Engagement, Staff Engagement, and Organizational Learning Capability were all significantly and positively related to Evaluation Use, there were differences on how these factors impacted the three different types of Evaluation Use. Consequently, this study provided clearer insights into which factors were more strongly connected to the different types of Evaluation Use, providing a focus for future evaluation capacity building interventions. Previous literature in this area has largely looked at each organizational characteristic in isolation and this study helped to identify which factors

emerged as important when the impact of multiple factors were assessed simultaneously. In addition, these differential relationships may be moderated by both Budget and Time Doing Evaluation, with nonprofits that have been conducting evaluation for more than seven years and/or nonprofits with budgets over \$1,000,000 having the resources and capability to create an infrastructure that enables greater use of evaluation for their nonprofit. This discussion section will address each of these areas, identify several limitations of the study, and provide insights into the implications of these findings for future research and action.

Evaluation Use

A review of the literature revealed an indistinct conceptualization of Evaluation Use. Very little attempt had been made to define use clearly and to emphasize the importance of distinguishing between use by local program staff, rather than high leveled decision makers, in addition to whether these results were used to benefit the organization itself or simply for accountability purposes. Multiple evaluation academics have emphasized the need for more systemic empirical work around evaluation use so as to enable comparability of organizations (Cousins et al., 2004; Leviton, 2003; Torres and Preskill, 2001; Weiss, 1988a, 1998). In an attempt to respond to the need expressed and move toward more empirical work in the nonprofit sector, this study set out to create an instrumental measure of Evaluation Use for the nonprofit sector, which differentiated between use for Accountability and use for Internal Learning.

Building on previous work by Carman (2005) and Fine et al. (1998), a more robust measure of Evaluation Use was developed, which conceptually, as well as statistically,

supported three different types of Evaluation Use: Accountability, Internal Learning, and Image Building. This study represents an initial testing and validation of use of evaluation for Accountability, for Internal Learning, and for Image Building in nonprofits. In the validation process, the exploratory factor analysis defined and then the measurement and theoretical path analysis models corroborated the existence of these three different types of Evaluation Use. Differentiating between the different types of Evaluation Use by the nonprofit allowed for comparability in terms of how the different independent variables impacted each type of Evaluation Use. While this study did not reveal dramatic differences between the factors associated with the different types of Evaluation Use, it did allow for some contrasts and emphasized that Organizational Learning Capability had a stronger impact on using evaluation results to improve Internal Learning rather than simply for Accountability or Image Building.

How Leadership Evaluation Characteristics, Stakeholder Engagement, and Organizational Learning Capability Impact Different Types of Evaluation Use

Correlational and multivariate analyses revealed that higher scores on Leadership Evaluation Characteristics, Internal Evaluator Engagement, and Internal Staff Engagement were positively related to higher levels of Evaluation Implementation and indirectly to Evaluation Use, while the dimensions underlying Organizational Learning Capability were directly related to Evaluation Use.

Leadership Evaluation Characteristics, such as the leader's evaluation beliefs and behavior, experience with evaluation, and perception of current evaluation practices,

repeatedly emerged as having the strongest relationship with Evaluation Implementation and Evaluation Use. This reemphasizes that having a leader who is supportive of the evaluation process makes evaluation activities more prevalent across the nonprofit, more likely to be implemented, and more likely to be used for multiple purposes. This supports the qualitative findings of Alaimo and Van Slyke (2006), who noted that organizations that used evaluation to inform their own programs and organizations had leaders who believed in the usefulness of the evaluation process and highlighted evaluation as a priority.

While the impact of Stakeholder Engagement on Evaluation Implementation and Use was not as strong as the impact of Leadership Evaluation Characteristics, the relationships were significant for two of the stakeholder groupings, Internal Evaluator Engagement and Staff Engagement, but not for Board Engagement or Executive Director Engagement. The lack of significance for Board and Executive Director Engagement could be a function of the lack of variability surrounding the board and executive director's lack of participation in evaluation activities that did not include decision-making, or a reflection of the fact that internal staff who are fully engaged in daily program activities are key to effective evaluation practices and use. These findings build on the earlier work of Fine, Thayer, and Coghlan (1998, 2000) and reiterate that, during the evaluation process, the stronger the involvement of staff internal to the nonprofit, such as evaluators or program staff, the greater the level of evaluation implementation and use of results. This suggests that nonprofits may benefit from engaging internal staff more thoroughly throughout the stages of the evaluation process, ensuring they are involved in decision-making stages in addition to the burdens of collecting data and preparing reports.

Finally, these analyses provide additional support for the research around Organizational Learning and reinforce the positive relationship between the presence of elements of a learning culture, such as managerial commitment, systems perspective, openness and experimentation, and knowledge transfer and integration and how nonprofits use evaluation to benefit their organizations (Botcheva, White, & Huffman, 2002; Jerez-Gómez et al., 2004; Preskill & Torres, 1999). Creating an organizational climate that supports learning enables evaluation results to be used effectively by that organization.

These general findings were supported by the overall path analysis models. In addition, running the path analysis models for each measure of Evaluation Use allowed for comparisons of these relationships across the different types of Evaluation Use. While the relationships between Leadership Evaluation Characteristics, Internal Evaluator Engagement, Staff Engagement, Organizational Learning Capability, and Implementation remained positive and significant for all three types of Evaluation Use, there were some discernable differences to interpret.

Leadership Evaluation Characteristics remained the most significant predictor of Implementation for use for Accountability, Image Building, and Internal Learning and Implementation continued to be strongly correlated with all three types of Evaluation Use and act as a mediator between Accountability, Image Building, and Internal Learning and Leadership Evaluator Characteristics, Internal Evaluator Engagement, and Staff Engagement. However, Organizational Learning Capability had a significantly stronger impact on whether a nonprofit used evaluation for Internal Learning, as opposed to Accountability or Image Building. This suggests that nonprofits who have leadership committed to evaluation and an

organizational culture that supports learning are more likely to use evaluation to benefit their organization rather than simply for external purposes such as Accountability or Image Building.

Moderating Effects of Budget and Time Doing Evaluation

While multiple nonprofit characteristics, such as Size, Budget, Age, Time Doing Evaluation, and Type of Nonprofit, were looked at in the analyses, only Time Doing Evaluation and Budget emerged as significant moderators of the relationships between Leadership Evaluation Characteristics, Evaluator Engagement, Staff Engagement, Organizational Learning Capability, and Accountability, Image Building, and Internal Learning. While the majority of these relationship differences were statistically significant, but minimal in meaning when graphed (Appendix H), there were some notable findings.

The effect of evaluation Implementation on Internal Learning was only significant for nonprofits with budgets of more than \$1,000,000 or for those that had been conducting evaluation for more than seven years, while this relationship was not significant for use of evaluation for Accountability or Image Building. Not surprisingly, this suggests that nonprofits that have the resources and/or capability to implement Evaluation Activities will be more likely to have the resources and/or capability to use evaluation results to benefit their nonprofit. In contrast, the effect of Organizational Learning Capability on Evaluation Use was only significant for nonprofits with budgets of less than \$1,000,000 or for those that had been doing evaluation for less than seven years. This suggests that Leadership Evaluation Characteristics may play a stronger role in the implementation and use of evaluation for

larger nonprofits with a background in evaluation, which could potentially be due to the more structured or hierarchical nature of the organization. Smaller, less experienced, nonprofits may continue to be more lateral in structure as the organizational climate is more governed by an overall focus on mission than dependent on characteristics of the executive director.

Limitations

The results of this study must be interpreted with caution due to several concerns primarily related to measurement issues that reduce external validity, but more importantly impact construct validity. The nonprofits that responded to this study were self-selected and part of a convenience sample based on relationships developed with three foundations in one geographic area. The sample of nonprofits may not be representative of the type of nonprofits in either North Carolina, or the U.S., and therefore these findings may not be applicable to other nonprofits in other regions across the U.S. or internationally.

The results of this study are further weakened by the social desirability bias of the self-report survey questions and the fact that the only respondent was the executive director of the nonprofit. As it is socially desirable for nonprofits to be seen using evaluation and for nonprofit leaders to be seen as supportive of evaluation practices, respondents may have overrated their self-reporting surrounding implementation, use of, and support for evaluation. This would result in an overestimation of both the importance of Leadership Evaluation Characteristics, in addition to an overestimation of Evaluation Implementation and Evaluation Use. More importantly, the use of only one respondent, the executive director, as a proxy for the entire nonprofit is problematic as this individual's responses become the level

of analysis for the constructs measured. As some of these constructs, such as Organizational Learning Capability, are measures that depict facets of the organizational climate, analysis based on a single respondent's self-reported answers may bias results.

There are several ways to address these limitations and attempt to improve measurement validity. While these improvements will be discussed fully in the next section on future research and action, they include using qualitative research methods, such as interviews, that can probe answers and elicit more specific information, surveying multiple respondents at different levels of the nonprofit, in addition to using archival data, such as nonprofit records, to support self-reported data.

Future Research and Action

This research builds on a new and growing body of empirical literature about nonprofits and program and organizational evaluation. While there are clearly limitations to the current study, it does make several important research contributions to the field. This is the first empirical study to be conducted with a large sample of nonprofits that uses multivariate analysis to simultaneously look at the interconnections between structural and organizational factors highlighted as predicting use of evaluation and organizational learning. In addition, one major contribution for this study relates to the development of clearer, more standardized measures for looking at the factors that facilitate use of evaluation by the nonprofit. Results from this study also provide an empirical basis for identifying some implications that could influence future evaluation capacity building initiatives.

To improve upon the existing study and increase measurement validity, different

approaches could be taken toward quantifying the constructs in future research. Leadership Evaluation Characteristics could be more appropriately estimated by a sample of employees, in addition to the Executive Director. Stakeholder Engagement could be better assessed through archival data, such as nonprofit records, or by asking the various stakeholders, such as the board, staff, and internal evaluators, about their level of participation. Organizational Learning Capability, as a measure of organizational climate, could be better estimated by multiple respondents at different levels of the nonprofit, in order to triangulate data. Implementation and Evaluation Use could be more accurately assessed through more objective ratings based on archival records or a stakeholder who may be better informed than the executive director. While resource intensive, these approaches may be more appropriate in terms of the level of analysis than simply using self-report data from a single respondent in a leadership position and consequently, could provide more objective ratings of the constructs.

The external generalizability of this study could also be improved by a more representative sample of nonprofits that would allow for comparisons across different types of nonprofits, paying particular attention to those nonprofits in the health and education sector that are under more pressure to be accountable. While this study found no discernable differences between organizational characteristics and the factors impacting use of evaluation, a more representative sample that resembled the state or U.S. nonprofit population characteristics may be able to differentiate between nonprofits that voluntarily, versus being obliged to, conduct and use evaluation for various purposes.

From a methodological standpoint, future empirical work in this area would benefit

from differentiating between the different types of evaluation use and ensuring that these measures are clearly outlined. Both qualitative and quantitative researchers should ensure that they are clearly defining use by whom (nonprofit, policymakers, public) and for what purpose (education, instrumental use, conceptual). This study has attempted to provide an instrumental measure of evaluation use for the nonprofit sector, which clearly differentiates between using evaluation for accountability, for image building, or for internal learning, but it is unknown whether these measures will be replicated in future studies. Future studies surrounding use of evaluation within the nonprofit sector may be able to use these measures to compare results in North Carolina to work in other geographic areas across the U.S. and even to contrast U.S. nonprofits with international organizations.

Future research should also look beyond the scope of this study toward the ultimate goal of using evaluation results to improve the overall effectiveness of nonprofits, making nonprofits more sustainable by enabling them to be more adaptable. While the assumptions underlying this study are that an organization that uses evaluation successfully is an innovating organization as it can clearly articulate inputs, activities, outputs, and outcomes and use these findings to adapt programs, improve client outcomes, and become more effective, there is no empirical data to support this. Just as nonprofits are being asked to demonstrate the effectiveness of their programs, those offering to build the evaluation capacity of nonprofits ought to ideally be able to demonstrate the effectiveness of their proposed interventions. Nonprofits, evaluation practitioners, and researchers could collaborate to collect data about the effectiveness of evaluation capacity building efforts, emphasizing how evaluation that feeds back into a nonprofit ultimately strengthens a

nonprofit and increases the nonprofit's sustainability and ability to make a valuable contribution to society.

The results of this study also have implications for future evaluation capacity building initiatives and where resources should be focused. At the organizational and program level, while the results for stakeholder engagement were not as strong as predicted, there was some support for the importance of staff engagement throughout the evaluation process, as opposed to board involvement, which was not significant. This highlights the relatively low impact of board involvement as compared to staff engagement, which could be explained by the lack of board involvement in the general daily operating of the nonprofit. This reiterates the importance of developing evaluation capacity initiatives that work at involving staff in the decision-making stages of evaluation, in addition to the more burdensome tasks of data collection and report writing. If staff feel they have a voice in the process and understand how results may benefit their organizations rather than be an added work burden, they may be more likely to collect quality and timely data that can be used to inform programs and practices and less likely to reject any negative findings.

Nonprofit leaders could develop motivational incentives that facilitate this staff involvement, providing rewards for staff that implement and document change based on evaluation results. In addition, any evaluation capacity building effort should recognize the tension within the nonprofit sector between conducting programs and investing resources to conduct and use evaluation effectively to improve their nonprofit. Interventions aimed at building evaluation capacity could therefore propose strategies that do not overburden staff and look at adapting evaluation tasks to fit more easily into the nonprofit's daily routine.

At the organizational level, assuming the strong relationship between a leadership supportive of evaluation and use of evaluation is not overestimated due to social responsibility bias, developers of evaluation capacity building initiatives should ensure they have buy-in from nonprofit leadership and work with nonprofit leaders to enhance their understanding around evaluation and how it can be used to benefit their nonprofit. If the nonprofit leadership is not supportive of evaluation and does not perceive its potential benefits, then they will not provide the necessary resources that will enable effective use of evaluation results.

Foundations could support this effort, by working with nonprofit leaders to develop mechanisms that would enhance organizational learning based on evaluation results. This could range from incentives similar to those proposed to enhance staff engagement, where foundations reward nonprofits that implement and document program and organizational changes based on results, to foundations working in partnership with nonprofits to allow for a stronger culture in which evaluation can be useful to nonprofits and seen less as an accountability mechanism. In addition, by allowing capacity building to be written into grant proposals, foundations could further acknowledge the importance of evaluation infrastructure to long-term sustainability of both the programs and nonprofits.

One further implication for developing evaluation capacity building initiatives is the relationship between having a climate that promotes organizational learning and using evaluation to benefit a nonprofit's programs and practices, rather than simply being used as a tool for accountability or image building. If the infrastructure and organizational climate are not considered in the design and implementation of evaluation capacity building initiatives,

then the outcomes may not be indicative of failure or success of the evaluation capacity building initiative, but rather of the current nonprofit learning culture capability. Dependent on the budget and amount of time doing evaluation, both leadership and organizational climate clearly impact how evaluation is used and consequently, will influence the effectiveness of any nonprofit evaluation capacity building initiatives.

This has connotations for both future research and action. From a research perspective, it will be critical to document the conditions under which evaluation capacity building efforts are deemed successful and contribute to the effectiveness of nonprofits. Systematic case studies, pre-post assessments, and the full range of research methodologies should be applied to determining what factors work in building evaluation capacity and impact nonprofit effectiveness. Again, this requires nonprofits, evaluation practitioners, and researchers to engage in a partnership to collect and disseminate information that can be used by the nonprofit sector to identify evidence-based practices in approaching evaluation capacity building. From an action perspective, this requires an understanding by those developing evaluation capacity building interventions that nonprofits vary in their willingness and capacity to engage in evaluation and that any intervention should be tailored to the level of commitment and capacity of the participants (Prochaska, Prochaska, & Levesque, 2001). Rather than setting standards for evaluation activities, nonprofits will need support in making steps toward their commitment to evaluation use. Effective evaluation capacity building interventions will therefore be those that address issues of leadership styles and organizational culture and recognize that, while there may be core components of evaluation capacity building, evaluation capacity building strategies may differ across

nonprofits varying in budget, size, resources, mission, and capacity.

VII. REFERENCES

- Alaimo, S. & Van Slyke, D.M. (2006). Examining the role of the leader in building organizational capacity for program evaluation in nonprofit human service organizations. Paper presented at the annual conference of the Association for Research on Nonprofit Organizations and Voluntary Action (ARNOVA), November 16-19, 2006, Chicago, Illinois.
- Alie, R.E. & Seita, J.R. (1997). Who's using evaluation and how? New study gives insight. *Nonprofit World*, 15(5) (September), 40-49.
- Alkin, M., Daillak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* Beverly Hills, CA: Sage.
- Allison, P.D. (2001). *Missing data*. Thousand Oaks, CA: Sage.
- Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Argyris, C. & Schön, D. A. (1978) *Organizational learning: A theory of action perspective*, Reading, MA Addison-Wesley.
- Argyris, C. & Schön, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison-Wesley.
- Aroian, L.A. (1944). The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics*, 18, 265-271.
- Avery, L.D. & VanTassel-Baska, J. (2001). Investigating the impact of gifted education evaluation at state and local levels: Problems with traction. *Journal for the Education of the Gifted*, 25(2), 153-176.
- Backer, T. E. (2000). *Strengthening nonprofits: Capacity-building and philanthropy*. Encino, CA: Human Interaction Research Institute.
- Balthasar, A. & Rieder, S. (2000). Learning from evaluations: Effects of the evaluation of the Swiss Energy 2000 programme. *Evaluation*, 6, 245-260.
- Bamberger, P. (1991). Reinventing innovation theory: Critical issues in the conceptualization, measurement and analysis of technological innovation. *Research in the Sociology of Organizations*, 9, 265-295.
- Baron, R.M. & Kenny, D.A. (1986). The moderator-mediator variable distinction in social

- psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Barrett, M. (2001). A stakeholder approach to responsiveness and accountability in non-profit organizations. *Social Policy Journal of New Zealand*, 17(December), 36-52.
- Bartholomew, L. K., Parcel, G. S., Kok, G., & Gottlieb, N. H. (2001). *Intervention Mapping: Designing Theory and Evidence Based Health Promotion Programs*. Mountain View, CA: Mayfield.
- Bauman, L.J., Stein, R.E.K., & Ireys, H.T. (1991). Reinventing fidelity: The transfer of social technology among settings. *American Journal of Community Psychology*, 19, 619-639.
- Beal, G. M., Rogers, E., & Bohlen, J. M. (1957). Validity of the concept of stages in the adoption process. *Rural Sociology*, 22(2), 166-168.
- Beardsley, R.M. & Miller, M.H. (2002). Revisioning the process: A case study in feminist program evaluation. *New Directions for Evaluation*, 96, 57-72.
- Bedell, J.R., Ward, J.C., Archer, R.P., & Kirk Stokes, M.. (1985). An empirical evaluation of a model of knowledge utilization. *Evaluation Review*, 9(2), 109-126.
- Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bentler, P. M. & Chou, C. P. (1987). Practical issues in structural equation modeling. *Sociological Methods & Research*, 16, 78-117.
- Bernstein, S.R. (1991). Contracted services: Issues for the nonprofit agency manager. *Nonprofit and Voluntary Sector Quarterly*, 20, 429-443.
- Bies, A. L. (2001). Accountability, organizational capacity, and continuous improvement: Findings from Minnesota's nonprofit sector. *New Directions for Philanthropic Fundraising*, 31, 51-80.
- Billitteri, T.J. 1998. Tracking the Effectiveness of Charities. *Chronicle of Philanthropy*, 10 (June 4), 35-37.
- Birleson, P. (1998). Learning organisations: A suitable model for improving mental health services? *The Australian and New Zealand Journal of Psychiatry*, 32(2), 214-222.
- Blakely, C. H., Emshoff, J. G., & Roitman, D. B. (1984). Implementing innovative programs in public sector organizations. In S. Oskamp (Ed.), *Applied Social Psychology Annual: Applications in organizational settings*, Vol. 5. Beverly Hills: Sage.

- BoardSource & INDEPENDENT SECTOR (2003; revised 2006). The Sarbanes-Oxley Act and implications for nonprofit organizations. Retrieved November 30, 2006 from: <http://www.boardsource.org/clientfiles/sarbanes-oxley.pdf>.
- Bohm, D. (1989). *On dialogue*. San Francisco: Harper.
- Botcheva, L., White, C. R., & Huffman, L. (2002). Learning culture and outcomes measurement practices in community agencies. *American Journal of Evaluation*, 23(4), 421-434.
- Bozzo, S. L. (2000). Evaluation resources for nonprofits: usefulness and applicability. *Nonprofit Management and Leadership*, 10(4), 463-472.
- Bozzo, S. L. (2002). Building evaluation capacity in the voluntary/nonprofit sector. *Canadian Journal of Program Evaluation*, 17, 75-92.
- Brandon, P.R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation*, 19(3), 325-337.
- Brody, E. & Cordes, J.J. (1999). Tax treatment of nonprofit organizations: A two-edged sword? In E.T. Boris, & C.E. Steurele, *Nonprofits and government: Collaboration and conflict* (pp.142-176). Washington, DC: The Urban Institute.
- Brown, D., Green, J., Hall, F., Rocchi, S., Rutter, P., & Dearing, A. (2000). Building a Better future: Innovation, Technology, and Sustainable Development. Geneva, Switzerland: World Business Council for Sustainable Development.
- Burke, C. B. (2001). Nonprofit history's new numbers (and the need for more). *Nonprofit and Voluntary Sector Quarterly*, 30(2), 174-203.
- Cadbury, A. (1998). The future for governance: the rules of the game. *Journal of General management*, 24(1), 1-14.
- Cadbury, A. (1999). What are the trends in corporate governance? How will they impact your company? *Long Range Planning*, 32(1), 12-19.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24(April), 263-274.
- Carman, J. G. (2001). Community foundations: A growing resource for community development. *Nonprofit Management and Leadership*, 12(1), 7-24.
- Carman, J. G. (2005). *Program evaluation use and practice: A study of nonprofit organizations in New York state*. University at Albany, State University of New York, New York.

- Carman, J. G., & Millesen, J. L. (2004). *Evaluation theory and practice: A report from the field*. Paper presented at the International Society for Third Sector Research, Toronto, Canada.
- Carver, J. & Carver, M.M. (1996). *Basic Principles of Policy Governance*. San Francisco, CA: Jossey-Bass.
- Caslyn, R., Tornatzky, L. G., & Dittmar, S. (1977). Incomplete evaluation of an innovation: The goal of attainment scaling. *Evaluation*, 128-130.
- Cattell, R.B. (1966). The scree test for a number of factors. *Multivariate Behavior Research*, 1, 245-276.
- Chelimsky, E. (1986). What have we learned about the politics of program evaluation? *Evaluation Practice*, 8, 5-21.
- Chen, H. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Christensen, R.A. & Ebrahim, A. (2006). How does accountability affect mission? The case of a nonprofit serving immigrants and refugees. *Nonprofit Management and Leadership*, 17(2), 195-209.
- Coleman, J. S., Katz, E., & Menzel, H. (1966). *Medical Innovation: A Diffusion Study*. New York: Bobbs-Merrill.
- Cook, J. B. (1988). Managing nonprofits of different sizes. In M. O'Neill & D. R. Young, (Eds.), *Educating managers of nonprofit organizations*, (pp.101-116). Westport, CT: Praeger.
- Cooley, W.W. (1983). Improving the performance of educational systems. *Educational Researcher*, 12(6), 4-12.
- Cousins, J.B. (1996). Consequences of researcher involvement in participatory evaluation. *Studies in Educational Evaluation*, 22(1), 3-27.
- Cousins, J.B. & Earl, L.M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis*, 14, 397-418.
- Cousins, J.B. & Earl, L.M. (1995). The case for participatory evaluation: Theory, research, and practice. In J.B. Cousins & L.M. Earl (Eds.), *Participatory Evaluation in Education: Studies in Evaluation Use and Organizational Learning* (pp.3-18). London: Falmer Press.
- Cousins, J.B, Goh, S.C., Clark, S. & Lee, L.E. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation*, 19(2), 99-141.

- Cousins, J.B. & Leithwood, K.A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56, 331-364.
- Cousins, J.B. & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 5-23.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Dale, N., Baker, A.J.L., & Racine, D. (2002). *Lessons learned: What the WAY program can teach us about program replication*. Washington, DC: American Youth Policy Forum.
- Davis, H.R. (1971). A checklist for change. In *A Manual for Research Utilization*. Washington, DC: Government Printing Office.
- De Vita, C. J. (1999). Nonprofits and devolution: What do we know? In E.T. Boris, & C.E. Steurele, *Nonprofits and government: Collaboration and conflict* (pp.213-245). Washington, DC: Urban Institute.
- DiBella, A. J., & Nevis, E. C. (1998). *How organizations learn: An integrated strategy for building learning capability*. California: Jossey-Bass.
- Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.
- DiMaggio, P. J. (1988). Nonprofit managers in different fields of service: Managerial tasks and management training. In M. O'Neill & D. R. Young, (Eds.), *Educating managers of nonprofit organizations*, (pp.51-69). Westport, CT: Praeger.
- Drasgow, F. (1988). Polychoric and polyserial correlations. In L. Kotz, & N.L. Johnson, (Eds.), *Encyclopedia of statistical sciences*, Vol.7 (pp.69-74). New York: Wiley.
- Drucker, P.F. (1990). *Managing the nonprofit organization: Practices and principles*. New York: Harper Collins.
- Ebrahim, A. (2005a). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly*, 34, 56-87.
- Ebrahim, A. (2005b). Nonprofit reporting in action: Practitioners and researchers assess the challenges of linking accountability to organizational learning. Paper presented at the annual conference of the Association for Research on Nonprofit Organizations and Voluntary Action (ARNOVA), November 17-19, 2005, Capital Hilton, Washington, D.C.

- Egan, T. M., Yang, B. & Bartlett, K. R. (2004). The effects of organizational learning culture and job satisfaction on motivation to transfer learning and turnover intention. *Human Resource Development Quarterly*, 15(3), 279-300.
- Ekvall G., Arvonen J., Waldenström-Lindström I. (1983). *Creative organizational climate: Construction and validation of a measuring instrument*. Stockholm: Swedish Council for Management and Work Life Issues.
- Ellinger, A.D., Yang, B., & Ellinger, A.E. (2000). Is the learning organization for real? Examining the impacts of the dimensions of the learning organization on organizational performance. *Proceedings of the Adult Education Research Conference Proceedings*, URL: <http://www.edst.educ.ubc.ca/aerc/2000/ellingeretal1-web.htm>.
- Emrick, J. A., Peterson, S. M., & Agarwala-Rogers, R. (1977). *Evaluation of the National Diffusion Network, Vol. 1: Findings and Recommendations*. Washington, DC: US Office of Education, Department of Health, Education, and Welfare.
- Fairweather, G. W. (1967). *Methods for experimental social innovation*. New York: Wiley.
- Fairweather, G.W. (1972). *Social change: The challenge to survival*. Morristown, NJ: General Learning Press.
- Fairweather, G.W. & Davidson, W. S. (1986). *An introduction to community experimentation: Theory, methods, and practice*. New York: Pergamon Press.
- Fairweather, G.W., Sanders, D., & Tornatzky, L. G. (1974). *Creating change in mental health organizations*. New York: Pergamon Press.
- Fairweather, G.W. & Tornatzky, L. G. (1977). *Experimental methods for social policy research*. Oxford: Pergamon Press.
- Fals-Borda, O. (1960). *Facts and theory of socio-cultural change in a rural system*. Monographias Sociologicas 2 bis, Universidad Nacional de Colombia, Bogota.
- Fetterman, D.M. (1994). Empowerment evaluation. American Evaluation Association Presidential Address. *Evaluation Practice*, 15, 1-15.
- Fetterman, D.M. (1996). Empowerment evaluation: An introduction to theory and practice. In D.M. Fetterman, S.J. Kaftarian, & A. Wandersman. (Eds.). (1996). *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability*. (pp.3-46). Thousand Oaks, CA: Sage Publications.
- Fine, A., & Jacobs, B. (2001). *Echoes from the field: Proven capacity-building principles for*

- nonprofits*. Washington, D.C.: Innovation Network, Inc.
- Fine, A. H., Thayer, C. E., & Coghlan, A. (1998). *Program evaluation practice in the nonprofit sector: A study funded by the aspen institute nonprofit sector research fund and the Robert Wood Johnson Foundation*. Washington, D.C.: Innovation Network, Inc.
- Fine, A. H., Thayer, C. E., & Coghlan, A. (2000). Program evaluation practice in the non-profit sector. *Nonprofit Management and Leadership*, 10(3), 331-339.
- Fixsen, D.L., Naoom, S.F., Blase, K.A., Friedman, R.M., and Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231)
- Fleishman, J. L. (1999). Public trust in not-for-profit organizations and the need for regulatory reform. In C.T. Clotfelter & T. Erlich, (Eds.), *Philanthropy and the nonprofit sector in a changing America* (pp.172-197). Bloomington, IN: Indiana University Press.
- Fliegel, F.C. & Kivlin, J. E. (1966). Attributes of innovations as factors in diffusion. *American Journal of Sociology*, 72(3): 234-248.
- Forbes, D.P. (1998). Measuring the unmeasurable: Empirical studies of nonprofit effectiveness from 1977-1997. *Nonprofit and Voluntary Sector Quarterly*, 27(2), 183-202.
- Frazier, P.A., Tix, A.P. & Barron, K.E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology*, 51,(1), 115-134.
- Freeman, R.E. (1984). *Strategic management: A stakeholder approach*. Boston: Pitman.
- Galaskiewicz, J. & Bielefeld, W. (1998). *Nonprofit organizations in the age of uncertainty: A study of organizational change*. New York, NY: Aldine De Gruyter.
- Garvin, D. A. (1993). Building a learning organization. *Harvard Business review*, 71(4), 78-91.
- Glosson, F.D. (2002). *Exploring the impact of the human dynamics program in a healthcare organization*. Unpublished doctoral dissertation, Cardinal Stritch University, Wisconsin.
- Goh, S. & Richards, G. (1997). Benchmarking the learning capacity of organizations. *European Management Journal*, 15(5), 575-83.

- Goodman, R. M., Steckler, A., & Kegler, M. C. (1996). Mobilizing organizations for health enhancement: Theories of organizational change. In K. Glanz et al., *Health Behavior and Health Education*. (2nd Ed.). San Francisco, CA: Jossey-Bass.
- Gray, S.T. (1998). *Evaluation with power - a new approach to organizational effectiveness, empowerment and excellence*. California: Jossey-Bass.
- Greenberg, B.S. (1964). Diffusion of news about the Kennedy assassination. *Public Opinion Quarterly*, 28, 225-232.
- Greene, J.C. (1987). Stakeholder participation in evaluation design: is it worth the effort? *Evaluation and Program Planning*, 10, 379-394.
- Greene, J.C. (1988a). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, 12(2), 91-116.
- Greene, J.C. (1988b). Communication of results and utilization in participatory program evaluation. *Evaluation and Program Planning*, 11(4), 341-351.
- Greenwood, M.R. (2001). Community as a stakeholder: Focusing on corporate social reporting. *Journal of Corporate Citizenship*, 4, 31-45.
- Gronjberg, K. A. (2001). The U.S. nonprofit human service sector: A creeping revolution. *Nonprofit and Voluntary Sector Quarterly*, 30(2), 276-297.
- Gronjberg, K. A. (1993). *Understanding nonprofit funding*. San Francisco, CA: Jossey Bass.
- Guba, E. G. & Lincoln, Y. S. (1989a). *Fourth Generation Evaluation*. Newbury Park, CA: Sage Publications.
- Guba, E. G. & Lincoln, Y. S. (1989b). The countenances of fourth-generation evaluation. In DJ Palumbo (ed.). *The politics of program evaluation*. Newbury Park, CA: Sage Publications.
- Guba, E.G. & Lincoln, Y.S. (1981). *Effective evaluation: Improving the usefulness of evaluation through naturalistic approaches*. San Francisco, CA: Jossey Bass.
- Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1999). *Multivariate data analysis*, 5th ed. New York: Prentice Hall.
- Hall, P.D. (2003). A solution is a product in search of a problem: A History of Foundations and Evaluation Research. Retrieved June 21, 2006 from <http://ksghome.harvard.edu/~phall/EVALUATION%20ESSAY.pdf>

- Hall, G. E. & Loucks, S. F. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal*, 14, 263-276.
- Hammack, D.C. (2001). Symposium: The quiet revolution in the nonprofit sector: Law, policy, and commerce. *Nonprofit and Voluntary Sector Quarterly*, 30(2), 157-173.
- Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation modeling*. Cary, NC: SAS Institute.
- Hatry, H.P. (1999). *Performance measurement: Getting results*. Washington, D.C.: Urban Institute.
- Hatry, H. van Houten, T., Plantz, M.C., & Greenway, M.T. (1997). *Measuring program outcomes: A practical approach*. Alexandria, VA: United Way of America.
- Havelock, R. G. (1973). *Planning for innovation through dissemination and utilization of knowledge*. Ann Arbor, MI: Institute for Social Research, University of Michigan.
- Henry, G. T. & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24(3), 293-314.
- House, E. R. (1981). Three perspectives on innovation: Technological, political, and cultural. In R. Lehming & M. Kane, (Eds) *Improving Schools: Using what we know*. (pp. 17-41). Beverly Hills, CA: Sage.
- House, E.R. (1993). *Professional evaluation: Social impact and political consequences*. Newbury Park, CA: Sage Publications.
- Howze, E. H. & Redman, L. J. (1992). The uses of theory in health advocacy: Policies and programs. *Health Education Quarterly*, 19(3), 368-383.
- Hughes, T.P. (1994). Technological Momentum. In M. R. Smith & L. Marx (Eds.). *Does technology drive history? The dilemma of technological determinism*. (pp. 101-113). Cambridge, MA: MIT Press.
- Hult, G.T. & Ferrell, O.C. (1997). Global organizational learning capacity in purchasing: construct and measurement. *Journal of Business Research*, 40, 97-111.
- INDEPENDENT SECTOR. (June, 2005). Panel on the Nonprofit Sector: Strengthening transparency, governance, accountability of charitable organizations. A final report to Congress and the nonprofit sector. Washington, D.C.
- Institute of Medicine. (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.

- Ismail, M. (2005). Creative climate and learning organization factors: their contribution towards innovation. *Leadership & Organization Development Journal*, 26(8), 639-654.
- Jeavons, T.H. (1994). Ethics in nonprofit management: Creating a culture of integrity. In R.D. Herman, (Ed.), *The Jossey-Bass Handbook of Nonprofit Leadership and Management* (pp.184-207). San Francisco, CA: Jossey-Bass.
- Jerez-Gómez, P., Céspedes-Lorente, J., & Valle-Cabrera, R. (2004). Organizational learning capability: A proposal of measurement. *Journal of Business Research*, 58, 715-725.
- Johnson, R.B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21, 93-110.
- Jones, R. B. (2003). Capacity building in human service organizations. *New Directions for Philanthropic Fundraising*, 40, 15-27.
- Jöreskog, K.G. & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications*, 2nd edition. Chicago: SPSS, Inc.
- Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kanter, R. M. (1988). When a thousand flowers bloom: Structural, collective, and social conditions for innovation in organization. In B. M. Staw & L. L. Cummings, *Research in Organizational Behavior*, 10, 169-211. New York, NY: JAI Press.
- Keith, T. Z. (2006). *Multiple regression and beyond*. Austin, Texas: Pearson.
- Kelly, J. G., Ryan, A, Altman, B. & Stelzner, S. (2000). Understanding and Changing Social Systems: An Ecological View. In J. Rappaport & E. Seidman (eds.) *Handbook of Community Psychology*. New York: Kluwer Academic/Plenum Publishers.
- Kelman, S. (2002). Contracting. In L.M. Salamon (Ed.). *The tools of government: A guide to new governance*. New York, NY: Oxford University Press.
- Kincaid, D. L. (2000). Social networks, ideation, and contraceptive behavior in Bangladesh: A longitudinal analysis. *Social Science and Medicine*, 50, 215-231.
- King, A.G. (1995). How to uncover jobs in nonprofit organizations. *Nonprofit World*, July/Aug, 55-57.
- Kinsey, D. J., & Raker, J. R. I. (2003). Capacity building: A primer. *New Directions for Philanthropic Fundraising*, 40, 5-14.

- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5-23.
- Kirsh, B., Krupa, T., Horgan, S., Kelly, D., & Carr, S. (2005). Making it better: Building evaluation capacity in community mental health. *Psychiatric Rehabilitation Journal*, 28(3), 234-241.
- Klein, K. J. & Sorra, J. S. (1996). The challenge of innovation implementation. *Academy of Management Review*, 21, 1055-1080.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd Edition). New York: Guildford Press.
- Kline, S. (1986). An overview of innovation. In R. Landau & N. Rosenberg (Eds.), *Positive Sum Strategy* (pp. 275-305). Washington, DC: National Academy Press.
- Kohl, J. W. (1966). Adoption stages and perceptions of characteristics of educational innovations. Ed.D. dissertation, University of Oregon, Eugene.
- Lampkin, L., Romeo, S., & Finnin, E. (2001). Introducing the nonprofit program classification system: The taxonomy we've been waiting for. *Nonprofit and Voluntary Sector Quarterly*, 30(4), 781-793.
- Larsen, J. K. & Agarwal-Rogers, R. (1977). *Re-invention of innovation: A study of community health centers*. Report, American Institute for Research in the Behavioral Sciences, Palo Alto, CA.
- Letts, C.W., Ryan, W., & Grossman, A. (1999). *High performance nonprofit organizations: Managing upstream for greater impact*. New York: John Wiley & Sons, Inc.
- Levin, B. (1987). The uses of research: A case study in research and policy. *The Canadian Journal of Program Evaluation*, 2(1), 43-55.
- Levin-Rozalis, M. & Rosenstein, B. (2003). A mentoring approach to the one-year evaluation course. *American Journal of Evaluation*, 24(2), 245-259.
- Leviton, L. C. (2003). Evaluation use: Advances, challenges, and applications. *American Journal of Evaluation*, 24(4), 525-535.
- Leviton, L. C. & Hughes, E. F. X. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation*, 5, 525-548.
- Light, P. C. (2000). *Making Nonprofits Work: A report on the tides of nonprofit management reform*. Washington, DC: Brookings Institute Press.

- Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Loucks, S. F., Newlove, B. W., & Hall, G. E. (1975). Measuring levels of use of the innovation: A manual for trainers, interviewers, and raters. University of Texas at Austin, R&D Center for Teacher Education, Austin, Texas.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 100, 107-120.
- MASTIC (Malaysian Science and Technology Information Center (1996). *1994 National survey of innovation in industry*. Ministry of Science, Technology and the Environment, Malaysia.
- Matlay, H. (2000). Organisational learning in small organizations: An empirical overview. *Education and Training*, 42(4-5), 202-210.
- Mayer, J. P. & Davidson, W. S. (2000). In J. Rappaport & E. Seidman (Eds.). *The handbook of community psychology*. NY: Plenum Press.
- McCaffrey, E. (2004). *Learning organizations in the WA public sector: An empirical study*. Curtin Business School, IPAA, Department of Premier & cabinet Fellowship Program Report.
- McGrath, R. G. (2001). Exploratory Learning, Innovative Capacity and Managerial Oversight. *The Academy of Management Journal*, 44(1), 118-131.
- McHargue, S. (2003). Learning for performance in nonprofit organizations. In Watkins, K. & Marsick, V. (Eds.). *Advances in developing human resources*, 5(2), p.196-204.
- McNelis, R. H., & Bickel, W. E. (1996). Building formal knowledge bases: Understanding evaluation use in the foundation community. *Evaluation Practice*, 17(1), 19-41.
- Merget, A.E. & Weaver, E.T. (1998). Organizational behavior and policy. In S.T. Gray, (1998), *Evaluation with power - a new approach to organizational effectiveness, empowerment and excellence*. (pp.63-73). California: Jossey-Bass,
- Monahan, J.I. & Scheirer, M. A. (1988). The role of linking agents in the diffusion of health promotion programs. *Health Education Quarterly*, 15, 417-433.
- Morley, E. Hatry, H. & Cowan, J. (2002). *Making use of outcome information for improving services: Recommendations for nonprofit organizations*. Washington, DC: Urban Institute.

- Morley, E., Vinson, E., & Hatry, H.P. (2001). *Outcome measurement in nonprofit organizations: Current practices and recommendations*. Waldorf, MD: Independent Sector.
- Murphy-Medley, D. (2005). Moving beyond accountability. Evaluation can be tool for organizational learning. *Philanthropy Journal*, (December).
- Nadler, D.A. (1983). Concepts for the management of organizational change. In J. R. Hackman, E.E. Lawler, III, & L.W. Porter (Eds). *Perspectives on behavior in organizations*. (pp. 551-561). New York: McGraw-Hill.
- Newcomer, K. E. (Ed.). (1997). *Using performance measurement to improve public and nonprofit programs*. California: Jossey-Bass.
- Newman, D.L., Brown, R.D., & Rivers, L.S. (1983). Locus of control and evaluation use: Does sense of control affect information needs and decision making? *Studies in Educational Evaluation*, 9, 77-88.
- OECD (1997). *OECD proposed guidelines for collecting and interpreting technological innovation data*. Paris, France: OECD/GD.
- Office of Management and Budget (1999). *Budget of the United States Government*. Washington, DC: U.S. Government Printing Office.
- Oldenburg, B. & Parcel, G. (2002). Diffusion of innovations. In K. Glanz et al., *Health Behavior and Health Education*, (pp.312-334). San Francisco, CA: Jossey-Bass.
- O'Sullivan, R.G. & O'Sullivan, J.M. (1998). Evaluation voices: promoting evaluation from within programs through collaboration. *Evaluation and Program Planning*, 14(3), 173-179.;
- Oswald, S.L., Mossholder, K.W., & Harris, S.G. (1994). Vision salience and strategic involvement: implications for psychological attachment to organization and job. *Strategic Management Journal*, 15, 477-489.
- Panel on Accountability and Governance in the Voluntary Sector. (February, 1999). Final report, Canada: Voluntary Sector Roundtable.
- Patton, M. Q. (1978). *Utilization-focused evaluation -the new century text*. Thousand Oaks, California: Sage Publications.
- Patton, M. Q. (1986). *Utilization-focused evaluation 2nd Edition*. Thousand Oaks, California: Sage Publications.
- Patton, M. Q. (1988). The evaluator's responsibility for utilization. *Evaluation Practice*,

- 9(2), 5-24.
- Patton, M. Q. (1997). *Utilization-focused evaluation, 3rd Edition*. Newbury Park, California: Sage Publications.
- Patton, M. Q. (2002). Feminist, yes, but is it evaluation? *New Directions for Evaluation*, 96, 97-108.
- Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan, N. J., French, B. D. & Blyth, D. A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. In Carol H. Weiss (Ed.), *Using Social Research in Public Policy Making* (pp.141-163). Lexington, New York: D.C. Heath.
- Plantz, M. C., Greenway, M. T., & Hendricks, M. (1997). Outcome measurement - showing results in the nonprofit sector. *New Directions for Evaluation*, 75, 15-30.
- Poole, D.L., Davis, J.K., Reisman, J., & Nelson, J.E. (2001). Improving the quality of outcome evaluation plans. *Nonprofit Management and Leadership*, 11(4), 405-421.
- Poole, D. L., Nelson, J., Carnahan, S., Chepenik, N. G., & Tubiak, C. (2000). Evaluating performance measurement systems in nonprofit agencies: The program accountability quality scale (PAQS). *American Journal of Evaluation*, 21(1), 15 -26.
- Porter, M. & Kramer, M.R. (1999). Philanthropy's new agenda: Creating value. *Harvard Business Review*, 77(November/December), 121-130.
- Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. *Evaluation Practice*, 18(3), 209-225.
- Preskill, H. & Torres, R. T. (1999). *Evaluative inquiry for learning in organizations*. California: Sage.
- Preskill, H. & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, 88, 25-37.
- Preskill, H., Zuckerman, B., & Matthews, B. (2003). An exploratory study of the factors that influence process use. *American Journal of Evaluation*, 24 (4), 423-442.
- Prochaska, J. O. (1979). *Systems of Psychotherapy: A Transtheoretical Analysis*. Pacific Grove, CA: Brooks-Cole.
- Prochaska, J. M., Prochaska, J. O., & Levesque, D. A. (2001). A transtheoretical approach to changing organizations. *Administrative and Policy in Mental Health*, 28(4), 247-261.
- Rappaport, J. M. (1977). *Community psychology: Values, research and action*. New York: Holt, Rhinehart & Winston.

- Robinson, T. & Cousins, J.B. (2004). Internal participatory evaluation as an organizational learning system: A longitudinal case study. *Studies in Educational Evaluation*, 31, 1-22.
- Rogers, E. M. (1971). *Diffusion of innovations*. New York: The Free Press.
- Rogers, E. M. (1983). *Diffusion of innovations*, (3rd Ed). New York: The Free Press.
- Rogers, E. M. (1995). *Diffusion of innovations*, (4th Ed). New York: The Free Press.
- Rogers, E. M. (2003). *Diffusion of innovations*, (5th Ed). New York: The Free Press.
- Rogers, E. M., Peterson, J. C., Cunningham-Sabo, L., & Davis, M. (2004). *Community participation and the utilization of health research*. Report, Prevention Research center, University of New Mexico, Albuquerque.
- Roome, N.J. (1998). *Sustainability strategies for industry*. San Francisco, CA: Island Press.
- Russ-Eft, D., Atwood, R., & Egherman, T. (2002). Use and non-use of evaluation results: Case study of environmental influences in the private sector. *American Journal of Evaluation*, 23, 19-31.
- Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organizations - a systematic approach to enhancing learning, performance and change*. Massachusetts: Perseus.
- Ryan, B. & Gross, N.C. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, 8, 15-24.
- Saidel, J.R. (1989). Dimensions of interdependence: The state and voluntary-sector relationship. *Nonprofit and Voluntary Sector Quarterly*, 18(4), 335-347.
- Salamon, L.A. (1995). *Partners in public service: Government-nonprofit relations in the modern welfare state*. Baltimore, MD: Johns Hopkins University Press.
- Salamon, L.A. (1999). *America's nonprofit sector: A primer* (2nd Ed). New York, NY: The Foundation Center.
- SAS Institute, Inc. (1989). *SAS/STAT users guide, version 6, fourth edition, volume 1*. Cary, NC: SAS Institute, Inc.
- Sawhill, J.C. & Williamson, D. (2001). Mission impossible? Measuring success in nonprofit organizations. *Nonprofit Management and Leadership*, 11(3), 371-386.
- Scribner, J., Cockrell, K., Cockrell, D., & Valentine, J. (1999). Creating professional communities in schools through organizational learning: An evaluation of a school

- improvement process. *Educational Administration Quarterly*, 35, 130-160.
- Seagal, S. & Horne, D. (1997). *Human dynamics: A new framework for understanding people and realizing the potential in our organizations*. Waltham, MA: Pegasus Communications.
- Seigart, D. & Brisolara, S. (2002). *Feminist evaluation - explorations and experiences*. California: Wiley.
- Senge, P. M. (1990a). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday Currency.
- Senge, P. M. (1990b). The leader's new work: Building learning organizations. *Sloan Management Review*, 32(1), 7-23.
- Senge, P. M., Roberts, C., Ross, R., Smith, B., & Kleiner, A. (1994). *The fifth discipline fieldbook: Strategies and tools for building a learning organization*. New York: Doubleday Currency.
- Shulha, L.M. & Cousins, J.B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195-208.
- Sielbeck-Bowen, K.A., Brisolara, S., Seigart, D., Tischler, C., & Whitmore, E. (2002a). Exploring feminist evaluation: the ground from which we rise. *New Directions for Evaluation*, 96, 3-8.
- Sielbeck-Bowen, K.A., Brisolara, S., Seigart, D., Tischler, C., & Whitmore, E. (2002b). Beginning the conversation. *New Directions for Evaluation*, 96, 109-113.
- Silber, N. I. (2001). *A corporate form of freedom: The mergence of the nonprofit sector*. Boulder, CO: Westview.
- Smith, S.R. (1993). The new politics of contracting: Citizenship and the nonprofit role. In H. Ingram & S.R. Smith (Eds.), *Public policy for democracy*. Washington, DC: The Brookings Institution.
- Smith, S.R. & Lipsky, M. (1993). *Nonprofits for hire: the welfare state in the age of contracting*. Cambridge, MA: Harvard University Press.
- Sobel, M.E. (1982). Asymptomatic intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp.290-312). San Francisco: Jossey-Bass.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7-25.

- Stoecker, R. (1997). The CDC model of urban development: A critique and an alternative. *Journal of Urban Affairs*, 19(1), 1-22.
- Stone, M. M. & Cutcher-Gershenfeld, S. (2001). Challenges of measuring performance in nonprofit organizations. In P. Flynn & V.A. Hodgkinson (Eds.), *Measuring the impact of the nonprofit sector*, (pp.33-57). New York, NY: Kluwer Academic/Plenum Publishers.
- Strichman, N. J. (2005). *Assessing nonprofits' organizational readiness for adaptive capacity*. Unpublished dissertation, University of Pittsburgh, PA.
- Stufflebeam, D.L. & Shinkfield, A.J. (1985). *Systematic Evaluation: A Self-instructional Guide to Theory and Practice*. Boston: Kluwer-Nijhoff.
- Sumariwalla, R.D. (1986, March). *Toward a national taxonomy of exempt entities*. Paper prepared for Independent Sector. Washington, D.C.
- Sundquist, J. L. (1968). *Politics and policy: The Eisenhower, Kennedy, and Johnson Years*. Washington, DC: The Brookings Institution.
- Taut, S. M. & Alkin, M.C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation*, 24(2), 213-226.
- Taylor, F.W. (1911). *The Principles of Scientific Management*. New York: Harper Bros.
- Taylor, M.E. & Sumariwalla, R.D. (1993), Evaluating Nonprofit Effectiveness: Overcoming the Barriers. In D. R. Young, R. M. Hollister, and V. A. Hodgkinson (Eds.), *Governing, Leading and Managing Nonprofit Organizations*, San Francisco: Jossey-Bass Publishers.
- Thayer, C.E. & Fine, A.H. (2000). Evaluation and outcome measurement in the non-profit sector: stakeholder participation. *Evaluation and Program Planning*, 24, 103-108.
- Tornatzky, L. G., Eveland, J. D., Boylan, M., Hetzner, W., Johnson, E., Roitman, D., & Schneider, J. (1983). *The process of technological innovation: Reviewing the literature*. Washington, DC: national Science Foundation.
- Tornatzky, L. G., Fergus, E. O., Avellar, J. W., Fairweather, G. W., & Fleischer, M. (1980). *Innovation and social process*. New York: Pergamon Press.
- Tornatzky, L. G. & Fleischer, M. (1990). *The processes of technological innovation*. NY: Lexington.

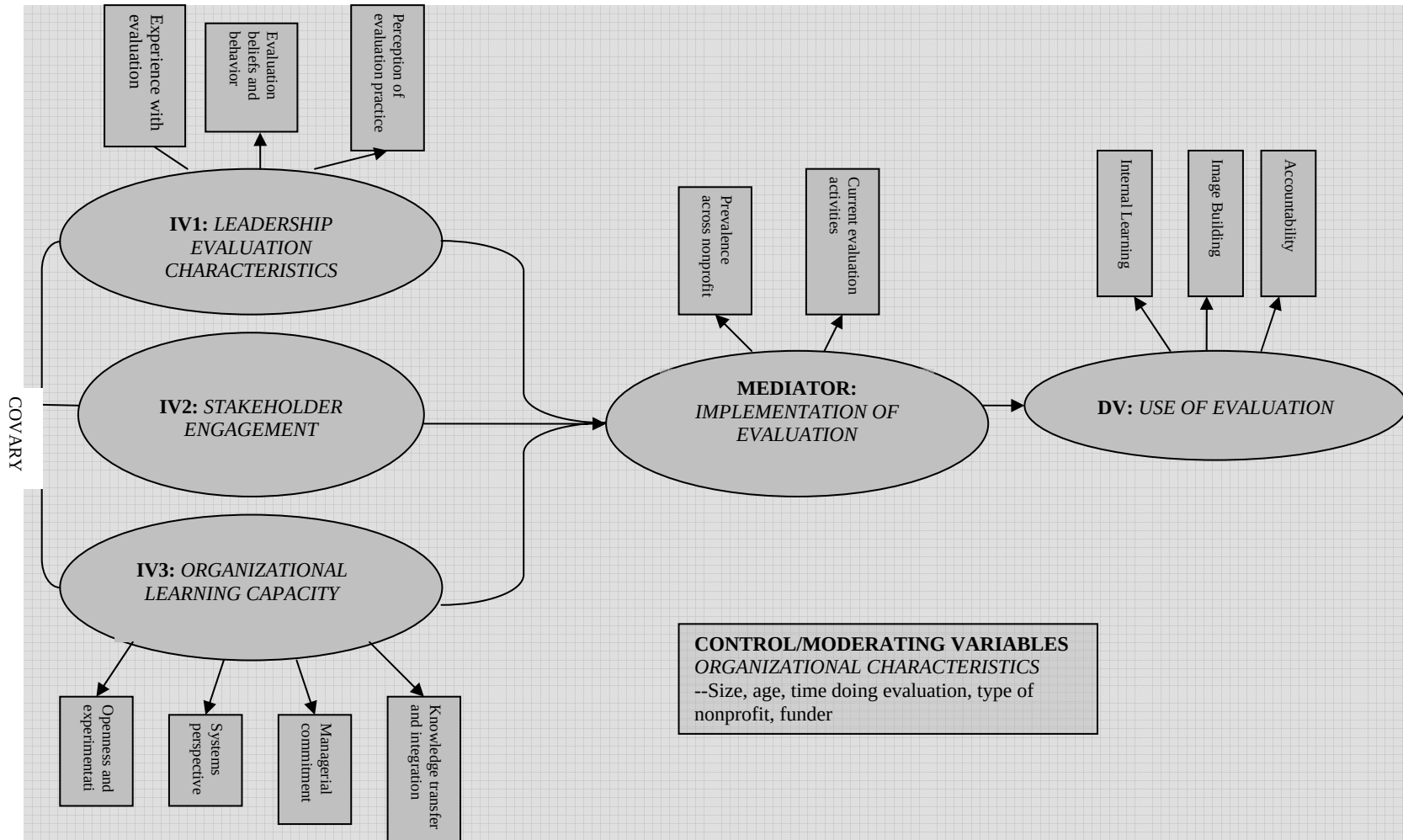
- Tornatzky, L. G. & Klein, K. (1982). Innovation characteristics and innovation adoption-implementation: A meta-analysis of findings. *IEEE Transactions on Engineering Management*, 29, 28-45.
- Torres, R. T., Preskill, H., & Piontek, M. E. (1996). *Evaluation strategies for communicating and reporting - enhancing learning in organizations*. California: Sage.
- Torres, R.T. & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, 22(3), 387-395.
- Truxillo, C. (2005). *Maximum likelihood parameter estimation with incomplete data*. SAS Users Group International Conference, Philadelphia, PA, April 10-13.
- Turnbull, B. (1999). The mediating effect of participation efficacy on evaluation use. *Evaluation and Program Planning*, 22, 131-140.
- Uebersax, J.S. (2000). Estimating a latent trait model by factor analysis of tetrachoric correlations. Statistical methods for Rater Agreement web site. Available at: <http://ourworld.compuserve.com/homepages/suebersax/irt.htm>. Accessed 03/7/2007.
- Urban Institute. (2003). National Center for Charitable Statistics. Core data files. Washington, DC: Urban Institute.
- Urban Institute. (2004). National Center for Charitable Statistics. Core data files. Washington, DC: Urban Institute.
- Urban Institute. (2006). National Center for Charitable Statistics. Core data files. Washington, DC: Urban Institute.
- U.S. General Accounting Office. (1998). *Program evaluation: Agencies challenged by new demand for information on program results. Report to the Committee on Governmental Affairs*, U.S. Senate (GAO-GGD-98-53).
- Vakola, M. (2000). Exploring the relationship between the use of evaluation in business process re-engineering and organizational learning and innovation. *Journal of Management Development*, 19, 812-835.
- Valovirta, V. (2002). Evaluation utilization as argumentation. *Evaluation*, 8, 60-80.
- Volkov, B. & King, J. (2005). A grounded checklist for implementing evaluation capacity building in organizations. Paper presented at the annual meeting of the American Evaluation Association in Toronto, Canada, October 2005.
- Ward, K.J. (2002). Reflections on a job done: Well? *New Directions for Evaluation*, 96, 41-56.

- Watkins, K. E. & Marsick, V. J. (1993). *Sculpting the learning organization: Lessons in the art and science of systemic change*. San Francisco, CA: Jossey-Bass.
- Watkins, K. E. & Marsick, V. J. (Eds.) (1996). *In action: creating the learning organization*. Virginia: American Society for Training and Development.
- Webber, A. M. (1999). Learning for a change. *Fast Company*, 24, 178-186.
- Weinstein, N. D. & Sandman, P. M. (2002). The precaution adoption process model. In K. Glanz et al., *Health Behavior and Health Education*, (pp.121-143). San Francisco, CA: Jossey-Bass.
- Weiss, C.H. (1981). Measuring the use of evaluation. In J. Ciaro (Ed.), *Utilizing evaluations* (pp.17-33). Beverly Hills, CA: Sage.
- Weiss, C.H. (1988a). Evaluation for Decisions: Is Anybody There? Does Anybody Care? *Evaluation Practice*, 9(1), 5-19.
- Weiss, C.H. (1988b). If program decisions hinged only on information: A response to Patton. *Evaluation Practice*, 9(3), 15-28.
- Weiss, C.H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21-33.
- Wholey, J. S. (1976). The Role of Evaluation and the Evaluator in Improving Public Programs: The Bad News, the Good News, and a Bicentennial Challenge. *Public Administration Review*, 36(6), (Nov. - Dec., 1976), 679-683
- Wholey, J. S. (1979). *Evaluation: Promise and performance*. Washington, DC: Urban Institute.
- Wholey, J. S. & Hatry, H. P (2001). The case for performance monitoring. *Public Administration Review*, 52(6), 604-610.
- Wholey, J. S., & Newcomer, K. E. (1997). Clarifying goals, reporting results. *New Directions for Evaluation*, 75, 91-98.
- Wholey, J. S., Scanlon, J.W., Duffy, H.G., Fukomoto, J.S., & Vogt, L.M. (1975). *Federal Evaluation Policy*. Washington, DC: The Urban Institute.
- Winter, S.G. & Sulanski, G. (2001). Replication as strategy. *Organization Science*, 12(6), 730-743.
- Wolch, J. (1990). *The shadow state: Government and voluntary sector in transition*. New York, NY: The Foundation Center.

- Worthen, B.R., Saunders, J.R., & Fitzpatrick, J.L. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*. New York: Longman.
- Yang, B. (2003). Identifying valid and reliable measures for dimensions of a learning culture. *Advances in Developing Human Resources* , 5(2), 152-162.
- Yang, B., Watkins, K. E., & Marsick, V.J. (1998). Examining construct validity of the Dimensions of the Learning Organization Questionnaire. In R. Torraco (Ed.), *Proceedings of the 1998 Annual Academy of Human resource development Conference*. Oak Brook, IL.
- Yang, B., Watkins, K. E., & Marsick, V.J. (2004). The construct of the learning organization: dimensions, measurement, and validation. *Human Resource Development Quarterly*, 15(1), 31-55.
- Zhang, D., Zhang, Z., & Yang, B. (2004). Learning organization in mainland China: empirical research on its application to Chinese state-owned enterprises. *International Journal of Training and Development*, 8(4), 258-273.

VIII: APPENDICES

APPENDIX A



Proposed model of the factors associated with the use of evaluation for organizational learning in North Carolina's nonprofit sector

APPENDIX B

List of variables from on-line survey**Leadership evaluation characteristics****Control variable**

25. How long have you worked in your current organization? ____ (years) (*control*)

--Experience with evaluation

26. As a leader in this organization...

- c. I have experience using evaluation in this and/or other organizations
- d. I have received training on how to use evaluation

--Evaluation beliefs and behavior

26. As a leader in this organization...

- a. I use systematically collected data to inform my decision-making
- b. I believe evaluation helps provide better programs, processes, services and products
- f. It is a priority for me that our organization invests resources in evaluation

-- Perception of current practice/Support for evaluation

26. As a leader in this organization...

- e. Currently available information tells us what we need to know about the effectiveness of our programs and processes.
- g. I feel our organization is doing an effective job in conducting and using evaluation

--Size, age, budget, organizational purpose

20. How many paid staff does your organization have (full-time equivalents)? _____

21. How many volunteer staff does your organization have (full-time equivalents)?

22. How long has your organization been operating? _____ years

23. What is your organization's current annual operating budget? (most recent completed fiscal year) *(This will be a space to fill in a number)*

18. Please select one category that you believe best matches the **purpose** of your organization (if this is not possible, please select the category which best matches your major programmatic activity):

- A Arts, Culture, and Humanities
- B Educational Institutions & Related activities
- C Envir. Quality, Protection & Beautification
- D Animal Related
- E Health – General and Rehabilitative
- F Mental Health, Crisis Intervention
- G Disease, Disorders, Medicinal Disciplines
- H Medical Research
- I Crime, Legal Related
- J Employment, Job Related
- K Food, Agriculture, and Nutrition
- L Housing, Shelter
- M Public Safety, Disaster Preparedness & Relief
- N Recreation, Sports, Leisure, Athletics
- O Youth Development
- P Human Services – Multipurpose and Other
- Q International, Foreign Affairs, National Security
- R Civil Rights, Social Action, Advocacy
- S Community Improvement, Capacity Building
- T Philanthropy, Voluntarism & Foundations
- U Science & Technology Research Institutes, Services
- V Social Science Research Institutes, Services
- W Public, Social Benefit: Multipurpose, Other
- X Religion Related, Spiritual Development
- Y Mutual/Membership Benefit Orgs., Other
- Z Other (please specify) _____

19. Could you briefly state your organization's mission?

--Barriers: Knowledge/skills/perceptions

14. In your organization, how important are the following barriers to using evaluation as an organizational learning tool? *(NB: scale will be provided above the question stems and respondents will click on box of choice; scale will read 1=not important, this is not a barrier, 2=moderately important, 3=very important, this prevents my organization from using evaluation for learning)*

- a. Not enough staff
- b. Not enough time
- c. Not enough evaluation expertise to conduct evaluation
- d. Staff lack knowledge of evaluation process
- e. Not enough funding
- f. Not enough training on how to use data, information, or evaluation tools
- g. Staff resistance to data collection
- h. Data collection or data management issues
- i. Lack of available tools to collect data/information
- j. Overly complex/time consuming tools to collect data/information
- k. Lack of *available* technical assistance
- l. Lack of *affordable* technical assistance
- m. Confidentiality issues
- n. Computer hardware problems
- o. Computer software problems
- p. Limited communication about reports or results within organization
- q. Problems with evaluation design
- r. Lack of *internal* leadership support (e.g. executive director, program directors)
- s. Lack of *board* support
- t. Funders require us to collect information that is not relevant to our own organizational needs
- u. Lack of incentives by funders (e.g. rewards, support for process, recognition)
- v. Lack of communication between external funders and organization
- w. Other issues (please specify):

--Evaluation activities

1. Does your organization currently practice any of the following management, oversight, or evaluation activities? (Please check ALL that apply to any of your organization's programs or services)

Conduct financial audits of your books
 Experience site visits by funders or regulatory agencies
 Acquire official licenses to operate programs (i.e., licenses for group homes, treatment programs)
 Participate in accreditation processes (i.e., Council on Accreditation, Council on Quality & Leadership, Joint Committee on Accreditation of Healthcare Organizations)

Conduct performance reviews and evaluations of program staff
 Review program documentation (i.e., participant or client records, case notes)
 Conduct firsthand observations of program activities
 Conduct formal program evaluations of your programs

Establish performance targets or program goals (i.e., serve 1000 people, 80% complete training)
 Monitor program implementation (i.e., to make sure programs are being delivered as intended)
 Assess whether you are meeting program goals and objectives
 Assess whether you are meeting organizational goals and objectives
 Engage in formal strategic planning processes

Produce annual reports
 Produce reports for funders about program activities
 Produce reports for funders about financial expenditures
 Produce reports for the board of directors

Use a performance measurement system (i.e., the United Way's outcome measurement system)
 Design program "logic models" (i.e., create diagrams linking inputs, processes & outcomes)
 Use a "balanced scorecard" management system (i.e., the system created by Kaplan & Norton)
 Use other management, evaluation and monitoring tools (please specify in the space provided)

2. What types of data or information has your organization collected in the last year? (please check ALL that apply to any of your organization's programs)

The number of people or organizations you serve
 The demographics of the people or organizations you served (i.e., age, ethnicity, income)
 Information about program expenditures (i.e., how much money you spend)
 Information about other resource expenditures (i.e., staff & volunteer time, equipment & supplies)

Information about consumer or participant satisfaction
 Information about best practices or benchmarks set by others in your field
 Information about program activities or outputs (i.e., hrs of counseling, # of referrals, hrs of technical assistance)
 Information about program outcomes or program results (i.e., changes in participant knowledge, changes in participant behaviors, improved conditions as a result of program activities)
 Narrative or anecdotal data (i.e., testimonials, stories about program participants)
 Control or comparison data (i.e., data from people you do not serve, to make comparisons)
 Other types of data or information: (please specify)

3. How does your organization collect data or information? (please check ALL that apply to any of your organization's programs)

Written data collection tools, including (please check all that apply)

Intake forms
 Minutes from meetings
 Exit surveys
 Activity logs
 Pre & post tests
 Case notes
 Other written tools (please specify)

Standardized tests or other standardized instruments
 Handheld computer systems/personal digital assistants (PDAs)
 Face-to-face interviews with program participants
 Telephone surveys of program participants
 Mail surveys of program participants
 Focus groups of program participants
 Observe and record program activities
 Other data collection methods (please specify)

6. Who has primary responsibility for conducting the management, oversight, or evaluation activities? (please select only one)

Internal/executive management staff
 Internal, evaluation staff
 Other internal staff (please specify)
 Board members or board committees
 Students (i.e. interns, class projects)
 External evaluator
 External agency or funder (please specify)
 Volunteers
 Other (please specify)

--Prevalence across organization

4. How would you characterize your organization's current overall approach to evaluation?
(please check only one statement)

We do not evaluate any of our programs and organizational activities

We evaluate a few selected programs and organizational activities (i.e. when required by a funder).

We evaluate more than half of our programs and organizational activities.

We evaluate all/almost all of our programs and organizational activities.

--Staff/stakeholder engagement

7. Please indicate all the stakeholders who participate in the following steps of the evaluation process: (please check ALL that apply): *If you do not do one of these activities, please check not applicable (N/A)*

Internal, executive/management staff
Internal, evaluation staff
Other internal staff
Board members or board committees
Students/volunteers (i.e., interns, class projects)
External evaluator
External agency or funder
Program participants
Other (please specify)
N/A We do not do this

- a. Decision-making regarding evaluation planning
- b. Design of the evaluation (conceptualization and deciding what questions to ask)
- c. Judging the appropriateness and quality of the evaluation's methodology
- d. Collecting data about programs or practices
- e. Interpreting the meaning of the evaluation data
- f. Writing the report
- g. Reading and using evaluation results
- g. Deciding how evaluation results are used by the organization

--Tool access, functionality, use

9b. How have your evaluation practices changed over the last few years (this could include resources needed, increased demand for accountability, improvements in perception of evaluation)? (*open-ended question*)

9c. Where do you currently go to get information about how to conduct and/or use evaluation? (if there is a URL, please include this) (*open-ended question*)

This will also be measured more as a control variable: using barrier information from q14

14. In your organization, how important are the following barriers to using evaluation as an organizational learning tool? (*NB: scale will be provided above the question stems and respondents will click on box of choice; scale will read 1=not important, this is not a barrier, 2=moderately important, 3=very important, this prevents my organization from using evaluation for learning*)

- a) Not enough staff
- b) Not enough time
- c) Not enough evaluation expertise to conduct evaluation
- d) Staff lack knowledge of evaluation process
- e) Not enough funding
- f) Not enough training on how to use data, information, or evaluation tools
- g) Staff resistance to data collection
- h) Data collection or data management issues
- i) Lack of available tools to collect data/information
- j) Overly complex/time consuming tools to collect data/information
- k) Lack of *available* technical assistance
- l) Lack of *affordable* technical assistance
- m) Confidentiality issues
- n) Computer hardware problems
- o) Computer software problems
- p) Limited communication about reports or results within organization
- q) Problems with evaluation design
- r) Lack of *internal* leadership support (e.g. executive director, program directors)
- s) Lack of *board* support
- t) Funders require us to collect information that is not relevant to our own organizational needs
- u) Lack of incentives by funders (e.g. rewards, support for process, recognition)
- v) Lack of communication between external funders and organization
- w) Other issues (please specify):

--Time

5. How long have you been conducting evaluation at your nonprofit organization? ____years

ORGANIZATIONAL LEARNING CAPABILITY (Likert scale, 7 point; 4 dimensions)

--Managerial commitment (MC)

MC1. The organization's leadership frequently involves staff in important decision-making processes.

MC2. Employee learning is considered more of an expense than an investment.

MC3. The organization's leadership looks favorably on carrying out changes in any area to adapt to and/or keep ahead of new environmental situations.

MC4. Employee learning capability is considered a key factor in this organization.

MC5. In this organization, innovative ideas that work are rewarded.

--Systems perspective (SP)

SP1. All employees have generalized knowledge regarding this organization's objectives.

SP2. All parts that make up this organization (departments, board, program staff, executive staff, volunteers, and other individuals) are well aware of how they contribute to achieving the overall objectives.

SP3. All parts that make up this organization are interconnected, working together in a coordinated fashion.

--Openness and experimentation (EX)

EX1. This organization promotes experimentation and innovation as a way of improving the work processes.

EX2. This organization follows up what other organizations in the nonprofit sector are doing, adopting those practices and techniques it believes to be useful and interesting.

EX3. Experiences and ideas provided by external sources (funders, technical assistance organizations, consultants, etc.) are considered a useful instrument for this organization's learning.

EX4. Part of this organization's culture is that employees can express their opinions and make suggestions regarding the procedures and methods in place for carrying out tasks.

--Knowledge transfer and integration (TR)

TR1. Errors and failures are always discussed and analyzed in this organization, on all levels.

TR2. Employees have the chance to talk among themselves about new ideas, programs, and activities that might be of use to the organization.

TR3. In this organization, teamwork is not the usual way to work.

TR4. The organization has instruments (manuals, databases, files, organizational routines, etc.) that allow what has been learnt in past situations to remain useful, although the employees are no longer the same.

Use of evaluation (frequency scale, 5 levels)

--Internal: organization practice changes based on results; outcomes modified;
programs/services improve

8. How often does your organization use evaluation for the following purposes? (never, seldom, about half the time, usually, always)

To help make changes in existing programs and/or services

To help make changes in organizational practices

For strategic planning purposes

To make decisions about fiscal allocations

To make decisions about staffing

To help develop new programs and/or services

To help us establish program goals or targets

To help us establish organizational goals or targets

To gain support for evaluation among staff or governing board

Other uses (please specify) _____

9. Could you provide an example of how your organization has used evaluation to inform organizational learning? (*open-ended*)

--External: feedback/communication between external funders and organization?

8. How often does your organization use evaluation for the following purposes? (never, seldom, about half the time, usually, always)

For outreach and public relations

To help us get new funding

To report to funders

To report to the board

Other uses (please specify) _____

APPENDIX C

E-mail cover letter script:

Subject header: Z. Smith Reynolds Foundation asks you to participate

Good morning,

Would you like to know how to ensure that your nonprofit organization's future evaluations are useful and lead to positive benefits for both your programs and organizations? So would we!

The Z. Smith Reynolds Foundation is cooperating with the Institute for Nonprofits at NC State University to conduct a survey of the foundations' grantees with the goal of identifying how evaluation can be more useful to North Carolina's nonprofit sector and lead to positive benefits for both programs and organizations. We hope that the information gathered in this study will help nonprofit leaders, funders, and support organizations to gather greater insights into how to develop systems that will facilitate using evaluation as an organizational learning tool.

Your participation will involve the completion of an on-line survey that should take between 15-20 minutes to complete. Deena Murphy-Medley, a Research Associate at the Institute for Nonprofits, will be contacting you via e-mail in the next day or so with a request that you complete an on-line survey. Please help us by participating in this on-line survey.

All responses will remain confidential. Please note that while the Z. Smith Reynolds Foundation supports this survey, your participation is entirely voluntary and will not impact the grant application process in any way. The results of the on-line survey may be published, but no identifying information will be used. Researchers at North Carolina State University will assign participating organizations a code and results will be analyzed using these codes so as not to identify any organizations or people individually. The Institute for Nonprofits will inform all participating organizations by e-mail when they have posted a general summary of the study results on the Institute's web site (<http://www.chass.ncsu.edu/nonprofit>).

If you have any questions concerning this on-line survey, please call Deena Murphy-Medley, Research Associate, at the Institute for Nonprofits, 919-513-7031, or e-mail at: deena_murphy-medley@ncsu.edu.

Sincerely,

Thomas W. Ross, Executive Director
Z. Smith Reynolds Foundation

APPENDIX D

SURVEY: Nonprofit organizations, evaluation practices, and organizational learning.
(Please note, survey is in an on-line format, notes about this design are in italics)

The Institute for Nonprofits at North Carolina State University, supported by The Community Foundation of Western North Carolina, the John Rex Endowment, and the Z. Smith Reynolds Foundation.

Would you like to know how to ensure that your organization's future evaluations are useful and lead to positive benefits for your organization? So would we!

The Institute for Nonprofits, in cooperation with The Community Foundation of Western North Carolina, the John Rex Endowment, and the Z. Smith Reynolds Foundation, is conducting a survey of the foundations' grantees to determine the factors that will make evaluation more useful to North Carolina's nonprofit sector. While we cannot initially promise you any direct benefit from your participation in this study, we hope that this experience will help you reflect on your organization's experiences, interest, and needs in the area of evaluation. We further hope that the information gathered in this study will help nonprofit leaders, funders, and support organizations to develop greater insights into how to develop systems that will facilitate using evaluation as an organizational learning tool.

Your participation is voluntary and you may choose not to answer any individual questions. All responses will remain confidential. Submitting the questionnaire indicates your consent to participate. Completing the questionnaire will take only about 15 minutes of your time. If you have any questions regarding this questionnaire or would like to be included in future work in this area, please contact Deena Murphy-Medley, Research Associate, at the Institute for Nonprofits, North Carolina State University, 919-513-0596, e-mail: deena_murphy-medley@ncsu.edu.

Your participation is greatly appreciated!

PLEASE ENTER THE NAME OF YOUR ORGANIZATION: _____
(submit button)

(this will take them to the survey main page and record their name for response record)

(Please note, this will remain confidential).

We would like to learn more about your organization's current evaluation practices. By evaluation, we mean the "process of systematic inquiry to provide information for decision-making about some object—a program, project, process, organization, system, or product." Please think about the typical practice of your organization over the last year and provide the following information about your organization's activities.

1. Does your organization currently practice any of the following management, oversight, or evaluation activities? (Please check ALL that apply to any of your organization's programs or services) (NB: activities will be in table and check boxes will be provided in on-line questionnaire)

Experience site visits by funders or regulatory agencies
 Acquire official licenses to operate programs (e.g., licenses for group homes, treatment programs)
 Participate in accreditation processes (e.g., Council on Accreditation, Council on Quality & Leadership, Joint Committee on Accreditation of Healthcare Organizations)
 Conduct performance reviews and evaluations of staff
 Conduct performance reviews and evaluations of board
 Review program documentation (e.g., participant or client records, case notes)
 Conduct firsthand observations of program activities
 Conduct formal program evaluations of your programs
 Establish performance targets or program goals (e.g., serve 1000 people, 80% complete training)
 Monitor program implementation (e.g., to make sure programs are being delivered as intended)
 Assess whether you are meeting *program* goals and objectives
 Assess whether you are meeting *organizational* goals and objectives
 Engage in formal strategic planning processes
 Produce annual reports
 Produce reports for funders about *program activities*
 Produce reports for funders about *financial expenditures*
 Produce reports for the board of directors
 Use a performance measurement system (e.g., the United Way's outcome measurement system)
 Design program "logic models" (e.g., create diagrams linking inputs, processes & outcomes)
 Use a "balanced scorecard" management system (e.g., the system created by Kaplan & Norton)
 Use other management, evaluation and monitoring tools (please specify in the space provided)

2. What types of data or information has your organization collected in the last year?

(Please check ALL that apply to any of your organization's programs) *(NB: activities will be in table and check boxes will be provided in on-line questionnaire)*

- The number of people or organizations you serve
- The demographics of the people or organizations you served (e.g., age, ethnicity, income, size, budget)
- Information about program expenditures (e.g., how much money you spend)
- Information about other resource expenditures (e.g., staff & volunteer time, equipment & supplies)
- Information about consumer or participant satisfaction
- Information about best practices or benchmarks set by others in your field
- Information about program activities or outputs (e.g., hours of counseling, number of referrals, hours of technical assistance)
- Information about program outcomes or program results (e.g., changes in participant knowledge, changes in participant behaviors, improved conditions as a result of program activities)
- Narrative or anecdotal data (e.g., testimonials, stories about program participants)
- Control or comparison data (e.g., data from people you do not serve, to make comparisons)
- Other types of data or information: (please specify)

3. How does your organization collect data or information? (Please check ALL that apply to any of your organization's programs) *(NB: activities will be in table and check boxes will be provided in on-line questionnaire)*

Written data collection tools, including (please check all that apply)

- Intake forms
- Minutes from meetings
- Exit surveys
- Activity logs
- Pre & post tests
- Case notes
- Other written tools (please specify) _____

- Standardized tests or other standardized instruments
- Handheld computer systems/personal digital assistants (PDAs)
- Face-to-face interviews with program participants
- Telephone surveys of program participants
- Mail surveys of program participants
- Formal focus groups of program participants
- Observe and record program activities
- Other data collection methods (please specify) _____

4. How would you characterize your organization's current overall approach to evaluation? (Please check only one statement) (NB: statements will be in table and check boxes will be provided in on-line questionnaire)

We do not evaluate any of our programs and organizational activities (*put skip in here to Q10, ignoring practices*)

We evaluate a few selected programs and organizational activities (i.e. when required by a funder).

We evaluate more than half of our programs and organizational activities.

We evaluate all/almost all of our programs and organizational activities.

5. How long has your nonprofit organization been conducting evaluation? (space provided)

_____ year(s) (*please round up to whole year*)

6. Who has primary responsibility for evaluation activities? (Please select only one)

Internal/executive management staff

Internal, evaluation staff

Other internal staff (please specify) _____

Board members or board committees

Students (i.e. interns, class projects)

External evaluators

External agency or funder (please specify) _____

Volunteers

Other (please specify) _____

Please use this space to elaborate on or clarify any of the above answers.

7. Please indicate all the stakeholders who participate in each of the following steps of the evaluation process: (please check ALL stakeholders that apply). If you do not do one of these steps, please check not applicable(N/A) in the far right column. (The activities will be listed down the side and the stakeholders (text upward) will be listed across the top with check boxes)

Executive management staff	
Internal, evaluation staff	
Other internal staff _____	
Members or board committees	
Students	
External evaluators	
Agency or funder _____	
Volunteers	
Program participants	
Other _____	
N/A	

- a. Decision-making regarding evaluation planning
- b. Design of the evaluation
- c. Judging the appropriateness and quality of the evaluation's methodology
- d. Collecting data about programs and practices
- e. Interpreting the meaning of the evaluation data
- f. Writing a report on evaluation findings
- g. Reading and using the evaluation results
- h.. Deciding how evaluation results are used by the organization

8. How often does your organization use evaluation for the following purposes: (NB: scale will be provided above the question stems and respondents will click on box of choice)

- o. For outreach
- p. For public relations
- q. To help make changes in existing programs and/or services
- r. To help make changes in organizational practices
- s. For strategic planning purposes
- t. To help us get new funding
- u. To make decisions about fiscal allocations
- v. To make decisions about staffing
- w. To help develop new programs and/or services
- x. To help us establish program goals or targets

- y. To help us establish organizational goals or targets
- z. To report to funders
- aa. To report to the board
- bb. To gain support for evaluation among staff or governing board
- cc. Other uses (please specify) : _____

9. a) Could you provide an example of how your organization has benefited from evaluation? (space provided)

b) How have your evaluation practices changed over the last few years (this could include resources needed, increased demand for accountability, improvements in perception of evaluation)

c) Where do you currently go to get information about how to conduct and/or use evaluation? (if there is a URL, please include this)

We would like to understand more about how your organization currently learns. Please respond to the following statements about the actual (not ideal) practices of your nonprofit organization, using the scale strongly disagree to strongly agree (in on-line version, scale will be provided above the question stems and respondents will click on box of choice)

		Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree
10a.	The organization's leadership frequently involves staff in important decision-making processes.							
10b.	Employee learning is considered more of an expense than an investment.							
10c.	The organization's leadership looks favorably on carrying out change in order to adapt to and/or keep ahead of new environmental situations.							
10d.	Employee learning capability is considered a key factor in this organization.							
10e.	In this organization, innovative ideas that work are rewarded.							
11a.	All employees have generalized knowledge regarding this organization's objectives.							
11b.	All parts that make up this organization (departments, board, program staff, executive staff, volunteers, and other individuals) are well aware of how they contribute to achieving the overall objectives.							
11c.	All parts that make up this organization are interconnected, working together in a coordinated fashion.							
12a.	This organization promotes experimentation and innovation as a way of improving the work processes.							
12b.	This organization follows up on what other organizations in the nonprofit sector are doing, adopting those practices and techniques it believes to be useful and interesting.							
12c.	Experiences and ideas provided by external sources (funders, technical assistance organizations, consultants, etc.) are considered a useful instrument for this organization's learning.							
12d.	Part of this organization's culture is that employees can express their opinions and make suggestions regarding the procedures and methods in place for carrying out tasks.							
13a.	Errors and failures are always discussed and analyzed in this organization, on all levels.							
13b.	Employees have the chance to talk among themselves about new ideas, programs, and activities that might be of use to the organization.							
13c.	In this organization, teamwork is not the usual way to work.							
13d.	The organization has instruments (manuals, databases, files, organizational routines, etc.) that allow what has been learned in past situations to remain useful, although the employees may no longer be the same.							

We would like to understand how best to facilitate nonprofit organizations using evaluation as an organizational learning tool. The following section asks you questions that will help nonprofit leaders, funders, and support organizations gain greater insights into how to develop systems that will facilitate using evaluation as an organizational learning tool.

14. In your organization, how important are the following barriers to using evaluation as an organizational learning tool? (NB: scale will be provided above the question stems and respondents will click on box of choice; scale will read 1=not important, this is not a barrier, 2=moderately important, 3=very important, this prevents my organization from using evaluation for learning)

- a. Not enough staff
- b. Not enough time
- c. Not enough evaluation expertise to conduct evaluation
- d. Staff lack knowledge of evaluation process
- e. Not enough funding
- f. Not enough training on how to use data, information, or evaluation tools
- g. Staff resistance to data collection
- h. Data collection or data management issues
- i. Lack of available tools to collect data/information
- j. Overly complex/time consuming tools to collect data/information
- k. Lack of *available* technical assistance
- l. Lack of *affordable* technical assistance
- m. Confidentiality issues
- n. Computer hardware problems
- o. Computer software problems
- p. Limited communication about reports or results within organization
- q. Problems with evaluation design
- r. Lack of *internal* leadership support (e.g. executive director, program directors)
- s. Lack of *board* support
- t. Funders require us to collect information that is not relevant to our own organizational needs
- u. Lack of incentives by funders (e.g. rewards, support for process, recognition)
- v. Lack of communication between external funders and staff in organization
- w. Other issues (please specify):

Please use this space to clarify or elaborate on barriers to using evaluation as an organizational learning tool:

15. In your organization, how useful are the following ways to communicate new knowledge and skills with staff at your organization? (NB: “useful” scale will be provided above the question stems and respondents will click on box of choice; scale will read 1=not useful, 2=moderately useful, 3=very useful)

- a. Conferences
- b. Workshops
- c. Newsletters
- d. Meetings
- e. Informal discussions
- f. Videos
- g. E-mails
- h. Web sites
- i. On-line training
- j. Written materials (e.g. training manuals)
- k. Other (please specify) _____

Please use this space to comment on other useful ways to communicate new knowledge and skills within your organization

16. How useful do you believe the following types of information, processes or activities would be to your organization? (NB: “useful” scale will be provided above the question stems and respondents will click on box of choice; scale will read 1=not useful, 2=moderately useful, 3=very useful)

- a. Techniques for coming to agreement about the primary goals and outcomes of your program(s) (e.g., logic models)
- b. How to identify and select appropriate indicators of important outcomes
- c. Planning and designing an evaluation plan that would answer your questions about program or organizational effectiveness
- d. Strategies for obtaining stakeholder buy-in regarding participation in evaluation
- e. Collecting monitoring data to determine whether your operations or programs are proceeding as planned
- f. Collecting (or finding) outcome data relevant to questions about effectiveness
- g. Analyzing and understanding the data you have collected (or have available)
- h. Reporting and defending your conclusions about your programs
- i. Techniques for sharing evaluation results with different audiences
- j. Other (please specify) _____

17. How useful do you believe the following types of resources would be to your organization?

(NB: “useful” scale will be provided above the question stems and respondents will click on box of choice; scale will read 1=not useful, 2=moderately useful, 3=very useful)

- a. An on-line searchable database on literature surrounding current evaluation practices
- b. On-line resources for conducting evaluation (e.g. case studies, methodologies, tools)
- c. A toolkit (written materials) for conducting evaluation
- d. Technical assistance (e.g. help from an outside agency) to design and maintain an evaluation system
- e. Agency specific training on how to do evaluation in your organization

- f. Organization-wide training on why and how to use evaluation to facilitate organizational learning
- g. Individual consultation that could guide your organization through the process of using evaluation for organizational learning
- h. Organized meetings and collaborations with similar nonprofits around evaluation issues (e.g., best use of time, staff, funds)
- i. Understanding how to pool resources with other nonprofits in order to obtain or share needed evaluation resources (e.g., hire an evaluator, purchase surveys, develop standardized indicators)
- j. Other (please specify): _____

What other types of evaluation information or resources do you believe would be useful to either your organization or the nonprofit sector in general?

Before you finish, please tell us about your nonprofit organization and yourself

18. Please select one category that you believe best matches the purpose of your organization (if this is not possible, please select the category which best matches your major programmatic activity):

- | | |
|---|--|
| A Arts, Culture, and Humanities | T Philanthropy, Voluntarism & Foundations |
| B Educational Institutions & Related Activities | U Science & Technology Research Institutes, Services |
| C Envir. Quality, Protection & Beautification | V Social Science Research Institutes, Services |
| D Animal Related | W Public, Social Benefit: Multipurpose, Other |
| E Health – General and Rehabilitative | X Religion Related, Spiritual Development |
| F Mental Health, Crisis Intervention | Y Mutual/Membership Benefit Orgs., Other |
| G Disease, Disorders, Medicinal Disciplines | Z Other (please specify) _____ |
| H Medical Research | |
| I Crime, Legal Related | |
| J Employment, Job Related | |
| K Food, Agriculture, and Nutrition | |
| L Housing, Shelter | |
| M Public Safety, Disaster Preparedness & Relief | |
| N Recreation, Sports, Leisure, Athletics | |
| O Youth Development | |
| P Human Services – Multipurpose and Other | |
| Q International, Foreign Affairs, National Security | |
| R Civil Rights, Social Action, Advocacy | |
| S Community Improvement, Capacity Building | |

19. Please briefly state your organization's mission:

20. How many paid staff does your organization have (full-time equivalents)? _____

21. How many volunteer staff does your organization have (full-time equivalents)? ____

22. How long has your organization been operating? _____ years

23. What is your organization's current annual operating budget? (most recent completed fiscal year) *(This will be a space to fill in a number)*

24. Which of the following positions best describes your role with this organization? (please check only ONE)

- Executive director
- Internal evaluator
- Assistant/deputy director
- External evaluator
- Chief financial officer
- Board member
- Senior staff member
- Other (please specify) _____

25. How long have you worked with this organization? ____ (years) *(space will be provided)*

Please respond to the following statements about yourself, using the scale strongly disagree to strongly agree *(scale will be provided above the question stems and respondents will click on box of choice)*

26. As a leader in this organization...

- a. I use systematically collected data to inform my decision-making
- b. I believe evaluation helps provide better programs, processes, services, and products
- c. I have experience using evaluation in this and/or other organizations
- d. I have received training on how to use evaluation
- e. Currently available information tells us what we need to know about the effectiveness of our programs and processes.
- f. It is a priority for me that our organization invests resources in evaluation
- g. I feel our organization is doing an effective job in conducting and using evaluation

Please use the following space to underline anything about using evaluation practices, use, funding, requirements or otherwise that you feel has not been addressed by this survey:

Thank you for completing this survey! Please click the Submit button to complete the survey

APPENDIX E

Table 27

Frequencies and Percentages for All Categorical Variables

Variable	Value	Frequency	Percent
Foundation	1	177	68
	2	70	27
	3	11	4
NTEE	Arts	6	2
	Education	21	8
	Environment	43	17
	Health	15	6
	Human	82	32
	International	1	<1
	Public	59	23
	Religion	1	<1
	Unknown	30	12
Evaluation Prevalence	None	2	<1
	Few	40	16
	Most	56	22
	All	160	62

APPENDIX F

Table 28

Descriptive Statistics for Numeric Variables and Scales

Variables/Scales	N	Mean	Median	St Dev	Min	Max	Kurtosis	Skew
<i>Organizational characteristics</i>								
Staff Size	255	39	11	60	1	430	9.8	2.9
Organization age	256	21	15	10	1	149	9.5	2.8
Budget	246	178,444	432,000	32,500	10,000	25,000,000	9.8	2.9
Evaluation years	245	10	7	8	1	56	5.2	1.9
<i>Implementation</i>								
Evaluation activities	258	6	6	2	0	11	-0.5	-0.1
<i>Stakeholder Engagement</i>								
Internal evaluator	258	3	2	3	0	8	-1.2	0.5
Internal staff	258	2	0	3	0	8	-0.7	0.9
Board	258	3	3	2	0	8	-1.1	0.1
Executive Director	258	7	7	2	0	8	2.7	-1.8

Table 28 (continued)

Variables/Scales	N	Mean	Median	St Dev	Min	Max	Kurtosis	Skew
<i>Organizational Learning Capability</i>								
Managerial commitment scale	258	6	6	1	1	7	6.2	-1.9
Systems perspective scale	258	6	6	1	1	7	5.2	-1.9
Openness and experimentation scale	258	6	6	1	2	7	3.9	-1.6
Knowledge transfer and integration scale	258	6	6	1	2	7	0.9	-0.9
<i>Leadership</i>								
Leadership evaluation characteristics scale	258	6	6	1	1	7	1.3	-0.8
<i>Use of evaluation</i>								
Internal learning scale	258	4	4	1	1	5	0.8	-0.8
Accountability scale	258	4	4	1	1	5	1.0	-0.8
Image building scale	258	3	3	1	1	5	-0.6	-0.2

Correlation Coefficients for Numeric Variables and Scales

[illegible]

APPENDIX H

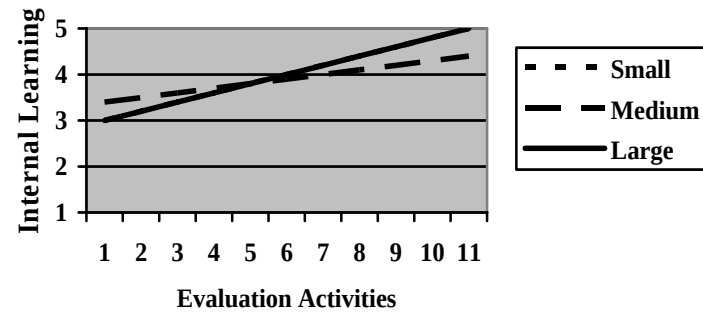


Figure 27. Effect of Evaluation Activities, conditional on Budget, on Internal Learning.

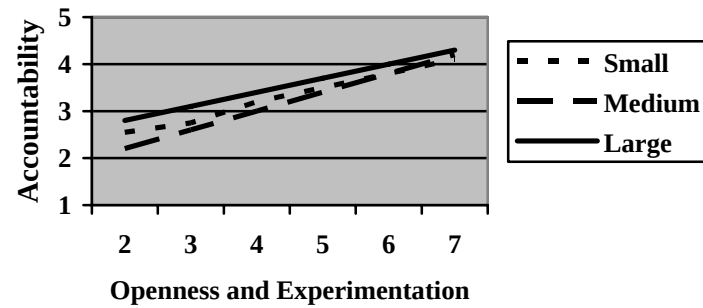


Figure 28. Effect of Openness and Experimentation, conditional on Budget, on Accountability.

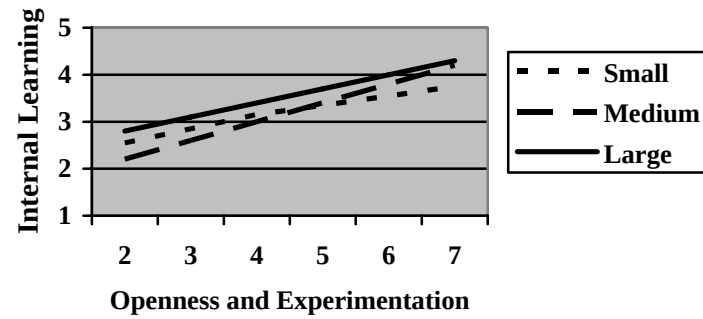


Figure 29. Effect of Openness and Experimentation, conditional on Budget, on Internal Learning.

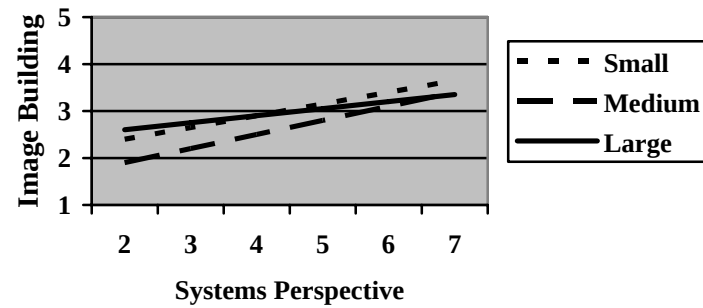


Figure 30. Effect of Systems Perspective, conditional on Budget, on Image Building.

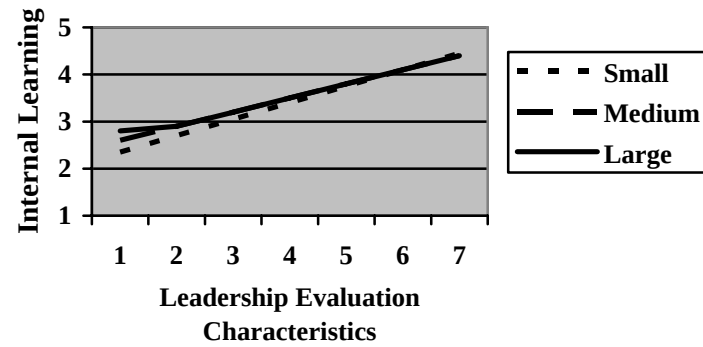


Figure 31. Effect of Leadership Evaluation Characteristics, conditional on Budget, on Internal Learning.

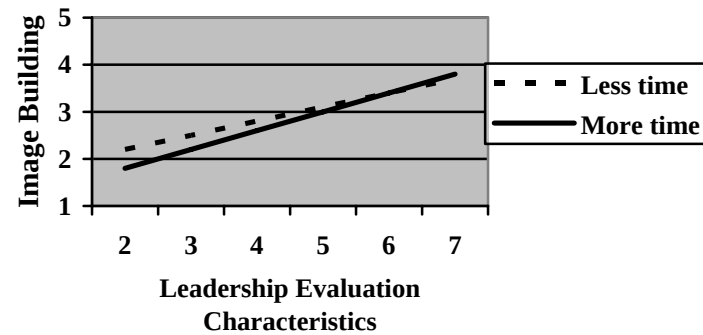


Figure 32. Effect of Leadership Evaluation Characteristics, conditional on Time Doing Evaluation, on Image Building.

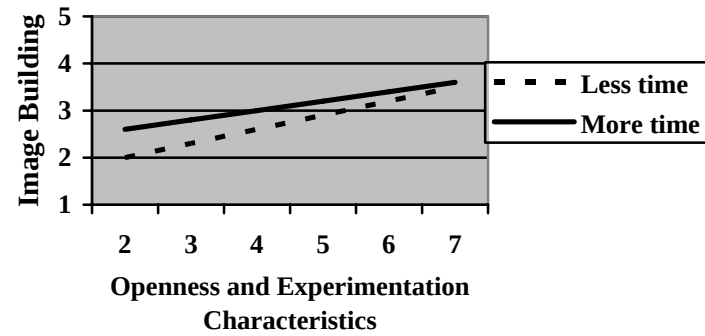


Figure 33. Effect of Openness and Experimentation, conditional on Time Doing Evaluation, on Image Building.

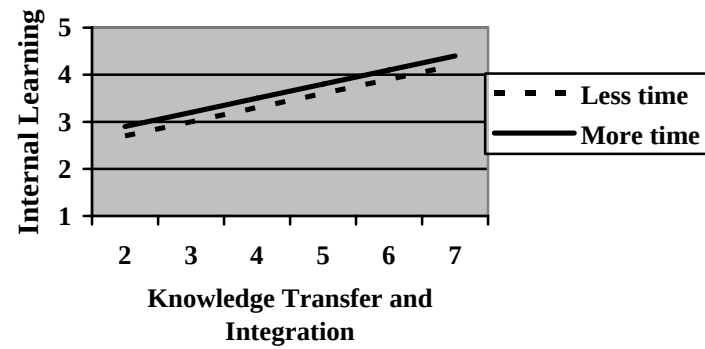


Figure 34. Effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation, on Internal Learning.

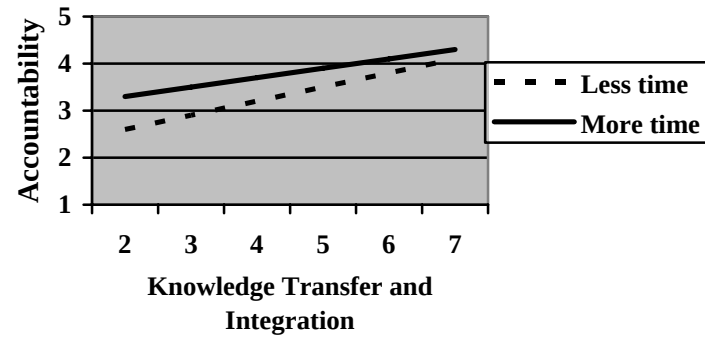


Figure 35. Effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation, on Accountability.

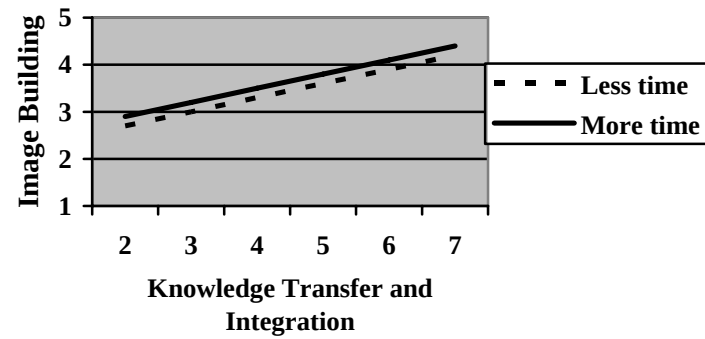


Figure 36. Effect of Knowledge Transfer and Integration, conditional on Time Doing Evaluation, on Image Building.

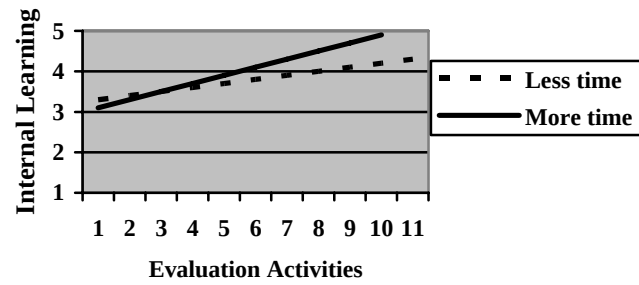


Figure 37. Effect of Evaluation Activities, conditional on Time Doing Evaluation, on Internal Learning.