

1-1-2014

Design and Validation of an Evaluation Checklist for Organizational Readiness for Evaluation Capacity Development

Connie F. Walker-Egea

University of South Florida, cwalkerpr@yahoo.com

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Curriculum and Instruction Commons](#)

Scholar Commons Citation

Walker-Egea, Connie F., "Design and Validation of an Evaluation Checklist for Organizational Readiness for Evaluation Capacity Development" (2014). *Graduate Theses and Dissertations*.
<http://scholarcommons.usf.edu/etd/5399>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Design and Validation of an Evaluation Checklist for Organizational Readiness for
Evaluation Capacity Development

by

Connie F. Walker-Egea

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Curriculum and Instruction
with a concentration in Measurement and Evaluation
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: Liliana Rodríguez-Campos, Ph.D.
John Ferron, Ph.D.
Robert Dedrick, Ph.D.
Roger Boothroyd, Ph.D.

Date of Approval:
October 9, 2014

Keywords: evaluation capacity building, nonprofit organization, internal evaluation, instrument
development, mnemonic device

Copyright © 2014, Connie F. Walker-Egea

DEDICATION

To Adriana, Julian, and Michelle

Anything is possible if you follow your dreams with all your heart.

ACKNOWLEDGMENTS

I am thankful to those who have been around and from the distance throughout all these years making this a wonderful journey of personal and professional growth. I feel very fortunate to had Dr. Liliana Rodríguez-Campos as my major professor and mentor. You were exceptionally supportive since I started the doctorate. Your dedication and guidance motivated me to work hard to be my best and reach my goals. As William Ward said, “The good teachers explain, the superior teachers demonstrate, the great teachers inspire.” Thank you for sharing with me the experience, expertise, and passion you have for what you do. I am also grateful for your confidence in me, and for the many opportunities you gave me to learn about evaluation, including working, publishing, and presenting papers with you.

I would also like to extent my gratitude to the members of my committee, Dr. John Ferron, Dr. Robert Dedrick, and Dr. Roger Boothroyd for your willingness to be part of it. Thank you for your kindness and for always being available for helping me to find answers to my questions and concerns even those not related to the dissertation. It was an honor to learn from you all and feel supported in many ways.

My special thanks to my friends and “dissertation buddies” Angie, Walter, and Vanessa for being with me in this adventure and for your unconditional help every time I needed. It was great to have people like you who understand what we have to go through; that made everything easier. My gratitude also goes to Heather Scott and Tyler Hicks for assisting me proofreading some of my work, and to those who participated in the different phases of the study for making this research possible.

I wish also to say a special thank you to my family who are an essential part of my life. You have been great in believing in everything I do. Thank you all for your infinite love and for everything you have done through my whole life.

Lastly, an immense thanks to God for everything you do every day of my life, for your unending blessings, for the people around me who make my life more meaningful, and for always being my source of inspiration. You made this dream possible!

TABLE OF CONTENTS

List of Tables	iv
Abstract	v
Chapter I: Introduction.....	1
Statement of the Problem.....	6
Purpose of the Study	8
Research Questions	9
Limitations	10
Overview of the Study	10
Organization of the Study	11
Definitions of Terms	11
Chapter II: Literature Review	16
What is Evaluation?	16
Importance of Evaluation.....	17
Evaluation Culture	20
Evaluation Capacity Building/Development Definitions	23
Evaluation Capacity Development	24
Evaluation Capacity Building	25
Evaluation Capacity Building/Development Process	27
Evaluation Capacity Building/Development Components	30
Organizational Environment.....	31
Organizational Leadership Support	32
Knowledge/Skills Development	34
Resources	36
Program Theory	37
Demand for Evaluation	38
Communication.....	39
Policies and Procedures	39
External Support	40
Incentives	41
Feedback Mechanism.....	42
Evaluation Use	42
Evaluation Capacity Building/Development Models and Frameworks.....	43
Evaluation Capacity Development Group Toolkit	44
Grounded Framework for Evaluation Capacity Building.....	45
Multidisciplinary Model of Evaluation Capacity Building	46

Evaluation Capacity: A Model.....	46
An Evaluation Capacity Building Training Framework.....	47
The Collaborative Immersion Approach to ECB.....	48
Integrated Evaluation Capacity Building Model.....	49
Evaluation Capacity Building/Development in Nonprofit Organizations.....	50
Internal Evaluation.....	51
Evaluation Checklists.....	55
Evaluation Capacity Building/Development Checklists.....	58
Summary.....	59
Chapter III: Methods.....	61
Purpose of the Study.....	61
Research Questions.....	62
Research Design.....	63
Validation Process.....	64
Face and Content Validity.....	65
Consequential Validity.....	66
Phase 1: Design of the Organizational Readiness for Evaluation Capacity	
Development Checklist and Expert Review Form.....	67
Purpose.....	67
Procedure.....	67
Phase 2: Expert Review.....	69
Purpose.....	69
Participants.....	70
Procedure.....	70
Data Analysis.....	71
Phase 3: Pretesting.....	73
Purpose.....	73
Participants.....	73
Procedure.....	74
Data Analysis.....	74
Phase 4: Field Study.....	75
Purpose.....	75
Participants.....	75
Demographics.....	82
Procedure.....	84
Data Analysis.....	86
Protection of Human Subjects.....	87
Summary.....	88
Chapter IV: Results.....	89
Phase 1: Design of the Organizational Readiness for Evaluation Capacity	
Development Checklist.....	90
Phase 2: Expert Review.....	91
Phase 3: Pretesting.....	106
Phase 4: Field Study.....	114

Psychometric Properties of the ORECD Checklist	114
Descriptive Statistics and Internal Consistency	115
Correlations	120
Appropriateness of the ORECD Checklist for the Intended Uses	126
Potential Consequences of Using the ORECD Checklist	129
Final Adjustments to the ORECD Checklist	134
Summary of Results by Research Question	141
Chapter V: Discussion	143
Summary of the Study	143
Discussion of Findings	145
Face and Content Validity Evidences of the ORECD Checklist	146
Suitability of the ORECD Checklist for the Intended Users	148
Consequential Validity Evidence of the ORECD Checklist	149
Psychometric Properties of the ORECD Checklist	151
Internal Consistency	151
Correlations	154
Limitations of the Study	156
Recommendations for Future Research	157
Implications for Practice	158
Conclusions	159
References	160
Appendices	173
Appendix A: Initial Version of the ORECD Checklist	174
Appendix B: Expert Review Form	178
Appendix C: Professional Background Questionnaire	183
Appendix D: Pretesting Version of the ORECD Checklist	184
Appendix E: Questions for Pretesting	188
Appendix F: Field Study Version of the ORECD Checklist	189
Appendix G: Questions for Field Study	193
Appendix H: Nonprofit Organization Demographics	194
Appendix I: Field Study Interview Protocol	195
Appendix J: Final Version of the ORECD Checklist	196
Appendix K: Approval Letter from the Institutional Review Board	200

LIST OF TABLES

Table 1: Types of Checklists.....	56
Table 2: Literature Reviewed to Design the Components and Items of the Organizational Readiness for Evaluation Capacity Development Checklist.....	68
Table 3: Composition of the Samples Selected.....	78
Table 4: Demographics of Nonprofit Organization Participants	83
Table 5: Descriptive Statistics for Relevance and Clarity for Each Original Item	93
Table 6: Action Taken with the ORECD Checklist Items After the Expert Review	100
Table 7: New Items Added to the ORECD Checklist.....	105
Table 8: Action Taken with the ORECD Checklist Items After the Pretesting.....	109
Table 9: Psychometric Properties of the ORECD Checklist Components	115
Table 10: Psychometric Properties of the ORECD Checklist Items.....	116
Table 11: Correlations of the ORECD Checklist Components.....	121
Table 12: Amount of Correlations by Component According to Their Magnitude.....	122
Table 13: Intercorrelations of the ORECD Checklist Items	123
Table 14: Action Taken with the Items of the ORECD Checklist After the Field Study	135

ABSTRACT

Evaluation capacity development (ECD) has been acknowledged as a system of processes to help organizations achieve sustainable evaluation practice. Examining the existing evaluation capacity of an organization before starting an ECD process is necessary and will increase the possibilities of success, determined by the establishment or strengthening of an evaluation system into the organization. In response to this need, this study involved the designing of the Organizational Readiness for Evaluation Capacity Development (ORECD) checklist and its initial validation, using a mixed method research design. The study was conducted in four phases, including: (a) the design of the ORECD checklist based on a review of the literature; (b) a review of the ORECD checklist by five experts to obtain face and content validity evidences, with emphasis on relevance and clarity of the items and how well the items fit the corresponding component; (c) a pretesting about the appropriateness of the wording of the items and format of the ORECD checklist by a sample of doctoral graduate students with formal training in evaluation and professional evaluators; and (d) a field study with 32 nonprofit organizations to determine the utility and benefits of using the ORECD checklist and potential improvements to the instrument. This phase generated information about the psychometric properties as well as consequential validity evidence. Findings indicated that the ORECD checklist has great potential to determine the readiness of an organization to develop evaluation capacity, as demonstrated by the feedback received from various groups of participants, establishing face, content, and consequential validity. Results from the psychometric analysis showed correlations

that, for the most part, suggested that the components are measuring aspects of the same construct. In addition, the alpha for most of the components supported the reliability of the ORECD checklist. The two components with alphas close to but below .70 required modifications in order to improve their reliability. Also, it was necessary to modify or reword some of the items. Ongoing efforts should provide information about how the changes made to the ORECD checklist are working and additional validity evidences as the one that can be obtained through factor analysis. This will allow the exploration of the underlying structure of the ORECD checklist and its components. It is expected that the ORECD checklist can be a contribution to the body of literature about ECD helping to address organizational readiness in order to support and sustain the development of evaluation capacity within organizations.

CHAPTER I

INTRODUCTION

Demand for systematic program evaluation information designed to improve organizational effectiveness and learning has led to various efforts to develop the evaluation capacity of organizations (Behrens & Kelly, 2008; Carman, 2007; Forss, Kruse, Taut, & Tenden, 2006; General Accounting Office [GAO], 2003; Huffman, Thomas, Lawrenz, 2008; King & Volkov, 2005; Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012; McDonald, Rogers, & Kefford, 2003; Newcomer, Hatry, & Wholey, 2004; Satterlund, Treiber, Kipke, Kwon, & Cassady, 2013; Taylor-Powell & Boyd, 2008; Volkov, 2011). Evaluation capacity development (ECD) is defined as “the capacity of putting in place structures that support evaluation efforts within an organization” (ECDG, 2009a).

According to Trevisan (2002), the field of evaluation is “increasingly recognizing the importance of evaluation capacity for the promotion, conduct, and utilization of effective evaluation” (p. 303). Considering that many organizations lack the skills to conduct systematic evaluations (Gibbs, Napp, Jolly, Westover, & Uhl, 2002), evaluation capacity development is becoming more common for helping organizations to meet the demand for information on what works, accountability requirements, and internal needs such as program improvement and the desire to increase evaluation use and funding sources (GAO, 2003; García-Iriarte, Suárez-Balcázar, Taylor-Ritzler, & Luna, 2011; Huffman et al., 2008; Naccarella et al., 2007; Preskill & Boyle, 2008).

External stakeholders, school districts, foundations and other nonprofit funders, federal and state government agencies, private organizations, and the public-at-large are very interested in evidence of organization effectiveness (Forss et al., 2006; Huffman et al., 2008; King, 2002; Newcomer et al., 2004; Picciotto, 1998; Satterlund et al., 2013; Stevenson, Florin, Mills, & Andrade, 2002). For example, the *Government Performance and Results Act* passed in 1993 requires government agencies to prepare annual reports on their progress in meeting their performance goals in order to contribute to data and decision-making by Congress based on the evaluation and to improve the accountability of the government. A few years later, the United Way of America (1996) published the *Measuring Program Outcomes* to guide grantees on documenting inputs, activities, outputs, and outcomes for those they served. In 2007, the Performance Improvement Council, an initiative of the federal government, was put in place to improve the performance of federal programs. In addition, there are other stakeholders such as program staff who want to know about the execution of their programs, so that they can improve these programs and learn from the information they gather (Newcomer et al., 2004).

Despite all this growing interest in the practice of evaluation and the promotion of capacity building at the organizational level has been difficult, particularly, for funders and their nonprofit grantees, to keep up the expectations in terms of their capacity to design, conduct, and use evaluation (Behrens & Kelly, 2008; Sobek & Agius, 2007). In other words, there is usually a gap between the capacity of an organization to conduct evaluation and the expectations from the funder (Satterlund et al., 2013). Specifically, some of the major challenges for any organization in this regard are the establishment of an evaluation process and the sustainability of this process using the resources available (King & Volkov, 2005). Volkov and Baron (2011) note, “Even though evaluation’s role is paramount for ensuring program success, many

organizations, both for- and nonprofits, struggle to fund external evaluations” (p. 103).

Organizations can address this situation, first, by recognizing the need to increase their internal evaluation capacity and second, by developing evaluation capacity to be able to design, execute, and manage program evaluation (Satterlund et al., 2013; Volkov & Baron, 2011).

Russ-Eft and Preskill (2009) indicate that “for too long, organizations have neglected integrating evaluation into their work processes and activities” (pp. 1-2). Wisely (2002) indicates that foundations have been slow incorporating evaluation to examine the results of their actions; nevertheless, Braverman, Constantine, and Slater (2004) point out that foundations are increasing their practice of evaluation. Although some organizations are starting to take advantage of the potential of conducting evaluation, many do not always recognize its numerous benefits (Russ-Eft & Preskill, 2009). Helping an organization to consider “why, when, where, and how evaluation will be used leads individuals to think in an evaluative way about how jobs are performed, how services are delivered, and how well the organization is run” (Evaluation Capacity Development Group [ECDG], 2009a, What is Evaluation Capacity Development?,” para. 3).

Evaluation capacity building (ECB) is a recent conceptual development (King & Volkov, 2005) and has been recognized as “a multidimensional construct involving maximizing potential in different areas” (Kirsh et al., 2005, p. 235). According to King and Volkov (2005):

The goal of ECB is to strengthen and sustain effective program evaluation practices by increasing an organization’s capacity to: design, implement, and manage effective evaluation projects; access, build, and use evaluative knowledge and skills; cultivate a spirit of continuous organizational learning, improvement, and accountability; and create awareness and support for program evaluation and self-evaluation as a performance

improvement strategy in the internal and external environments in which they function.

(p. 11)

The ultimate goal of ECB is to achieve sustainable evaluation practice and the use of evaluation for decision making (Adams & Dickinson, 2010; Compton, 2009; Preskill & Boyle, 2008), meaning that, the organization is committed to incorporate “evaluation processes, systems, policies, and procedures that are self-renewing and evolving” (Preskill & Boyle, 2008, p. 454) as well as continuous learning about evaluation (Taylor-Ritzler, Suárez-Balcázar, García-Iriarte, Henry, & Balcázar, 2013). In addition to evaluation use, Bourgeois and Cousins (2013) recognize also organizational learning as a consequence of the ECB processes.

When evaluation capacity is deficient or nonexistent, organizations confront major constraints to conduct evaluation and respond to the demands of stakeholders (Huffman et al., 2008). Therefore, there are many specific reasons that make the development of evaluation capacity an essential piece of an organization. Among some of them are: (a) increases the efficiency and effectiveness of the operations of the organization, decreasing costs (ECDG, 2009b; Owen, 2003); (b) helps organizations improve levels of performance (Mackay, 1999) and their decision-making process (Boyle & Lemaire, 1999); (c) supports the existence of an evaluation culture in which individuals within the organization are able to understand the importance of and contribute to the evaluation, encouraging the meaningful use of its results (Owen, 2003); (d) makes available sustainable resources for producing evaluation (McDonald et al., 2003); (e) guides the efforts for quality improvement (ECDG, 2009b); and (f) promotes organizational learning (Bourgeois & Cousins, 2013; Hauge, 1998; Owen, 2003; Russ-Eft & Preskill, 2009).

The development of evaluation capacity varies depending on factors such as the availability of resources, the complexity of the organization, and the organizational capacity (i.e., capabilities, knowledge, and resources that nonprofits need in order to be effective). It is also strongly related to organizational maturity (Love, 1983). Piccioto (1998) indicates that if organizational capacity is not well established evaluations are simply paper exercises, and of no value. Thus, the development of evaluation capacity should be addressed jointly with the managerial processes of an organization to obtain successful results (Hauge, 1998).

Another challenge that organizations may face while developing evaluation capacity is the evolution of new ways of thinking about how to use evaluation and other empirical data in decision-making (Compton, 2009). As Forss et al. (2006) suggest, building evaluation capacity requires a consolidated effort by the people involved in it because it is not an automatic process but an organized implementation process that requires clear guidance and ongoing support. Also, ECD should take into consideration the needs of the personnel and provide direction to help them (Sonnichsen, 1999). “In developing ways to help organizations assess their work, examine what is working and what is not, and learn how to strengthen program activities and increase their impact, building the evaluation capacity of organizations can be of paramount significance” (King & Volkov, 2005, p. 15).

Evaluation capacity building is increasingly being acknowledge not only as a mechanism for providing accountability for resource use but also as an instrument of organizational learning (Hauge, 1998; Volkov, 2011) and change (Volkov, 2011). Preskill (2008) explains that “an organization’s ability to learn is a critical factor associated not only with survival but also with continued success” (p. 129). Indications that a learning environment is already in place is when an organization integrates evaluation capacity building as an institutional practice and priority

(Forss et al., 2006). The efforts to build evaluation capacity while promoting organizational learning usually focus on making changes within the organization and the way people think about the evaluation process (McDonald et al., 2003; Preskill, 2008). As Compton (2009) highlights, ECB is a type of organizational change.

It is important to note that, for practical purposes, the term evaluation capacity building has been used interchangeably with evaluation capacity development throughout the literature (Schaumburg-Müller, 1996) and in organization and agency documents and reports. Evaluation capacity building is the current terminology being used in North America (Levin-Rozalis, Rosenstein, & Cousins, 2009) and evaluation capacity development is the terminology used by the World Bank (Levin-Rozalis et al., 2009) and internationally (Stockdill, Baizerman, & Compton, 2002). Accordingly, the review of the literature for this study consists of both ECB and ECD. In addition, ECB and ECD acronyms will be used interchangeably throughout the study.

Statement of the Problem

Formal evaluations require systematic efforts characterized by methodological procedures. In the absence of an evaluation system, there is lack of information or evidence of excellent quality that can be used to help an organization accomplish its goals. This is the main reason to establish an evaluation system. It is worth noting that the evaluation system of an organization “needs to include a variety of evaluation activities that serve the needs of different organizational stakeholders, purposes, and uses” (Taut, 2007b, p. 57). ECD is an example of a system of guided processes and practices. It requires ongoing and active support from the people in charge of conducting the processes (e.g., organizational leadership) to incorporate it into the organizational culture.

Developing evaluation capacity is challenging and it includes a broad range of difficulties and issues (Huffman et al., 2008). Thus, to establish or strengthen an internal evaluation system through developing evaluation capacity it is necessary to identify first the existing capacity and the status of that system in the organization. Khan (1998) indicates that “No ECB support should take place without thorough analysis and examination of existing institutional capacities and their future operational sustainability” (p. 323). Because organizations find themselves at different stages of preparedness for ECB (Cohen, 2006), determining organizational readiness to develop evaluation capacity before engaging in this process is essential (García-Iriarte et al., 2011; Naccarella, Pirkis, Kohn, Morley, Burgess, & Blashki, 2007; Taut 2007b).

Although, there are other checklists that address the ECB/ECD topic, including *A Checklist for Building Organizational Evaluation Capacity* by Volkov and King (2007), the *Evaluation Capacity Development: A Diagnostic Guide and Action Framework* by the World Bank (Mackay, 1999), and the *Institutionalizing Evaluation Checklist* by Stufflebeam (2002), none of these checklists were designed to evaluate the readiness of an organization to embark in the development of evaluation capacity at any given point. Specifically, the checklist developed by Volkov and King was designed to provide a set of guidelines for organizational evaluation capacity building. The World Bank document focuses on the development of a country’s evaluation capacity, and was created to assist governments and development agencies which have decided to develop a national or sectorial evaluation system. Lastly, the checklist developed by Stufflebeam identifies steps that need to be taken in order to install or strengthen, assess, and maintain an evaluation unit supporting the development of evaluation capacity within an organization. Thus, the design of an evaluation checklist addressing organizational readiness for developing evaluation capacity provides organizations a dependable means to identify the

situation of their internal evaluation capacity and guide them in the establishment or strengthening of an internal evaluation system to move the ECD/ECB effort forward. The intention is to help support the institutionalization of evaluation into the organizations.

Purpose of the Study

This study was conducted for designing and validating a checklist intended to determine the readiness of an organization to develop evaluation capacity. This instrument provided nonprofit organizations (classified as 501(c)(3) public charities) guidelines containing a formal structure to establish the extent to which they are prepared for the development of evaluation capacity, enhancing objectivity and reproducibility of the assessment. The Organizational Readiness for Evaluation Capacity Development (ORECD) Checklist allows organizations, first, to identify their current situation to support the development of internal evaluation capacity; second, to guide the organization in recognizing which areas may be in need of improvement; and third, to determine the progress made by the organization toward readiness for developing evaluation capacity, by revisiting the ORECD checklist when necessary.

The structure of the ORECD checklist consist of various components (e.g., organizational environment, organizational leadership support, resources) identified in the literature, known to contribute to developing evaluation capacity within organizations. A set of items was developed to represent each of these components. The *Standards for Educational and Psychological Testing* [The Standards] (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999) and the steps for test construction (Crocker & Algina, 2008) were the main guidelines for designing and validating the ORECD checklist.

Research Questions

The following research questions guided this study:

1. To what extent does the Organizational Readiness for Evaluation Capacity Development Checklist integrate elements presented in the literature that support evaluation capacity development/building based on expert review?
 - a. What is the evidence for face validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
 - b. What is the evidence for content validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
2. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist suitable for the intended users as perceived by individuals with formal evaluation training?
3. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist appropriate for the intended uses as perceived by nonprofit organizations?
4. What are the potential positive and negative consequences of the Organization Readiness for Evaluation Capacity Development Checklist?
5. What are the psychometric properties of the Organizational Readiness for Evaluation Capacity Development Checklist in the field study?
 - a. What are the item-to-total correlations for each component?
 - b. What is the inter-item relationship for each component?
 - c. What is the relationship between the components of the checklist?
 - d. What is the relationship between all the items in the checklist?
 - e. What is the internal consistency for each component?

Limitations

The scope of this study was confined to the data collected in nonprofit organizations classified as 501(c)(3) public charities. Nonprofit organizations are just one possible scenario for the development of evaluation capacity. Therefore, the ORECD checklist must be used carefully with other types of organizations because the nature of nonprofit work (e.g., evaluation requirements from multiple funders), the organizational structure, and context could be different. ECD is context dependent, consequently, any measure developed to be used in one setting, perhaps needs to be adapted to be used in a different setting.

Overview of the Study

The study was conducted in four phases. Phase 1 comprised the design and construction of the ORECD checklist, based on a review of the literature. The Expert Review Form was also developed as part of this phase. The majority of the literature searched was restricted to publications from 2000 to 2014, including topics such as ECB/ECD at the individual and organizational level, evaluation, evaluation culture, internal evaluation, evaluation in nonprofit organizations, and evaluation checklists. Publications prior to 2000 included in the review are, for the most part, supporting literature in evaluation and measurement to provide a historical and methodological context, respectively. Only original articles and books published in English were included. The literature search included Educational Full Text and ERIC databases, as well as evaluation journal websites and Google Search. Additional articles and books identified through the references section of the articles initially reviewed were also included. Keywords used for searching the literature were: checklist, evaluation, evaluation capacity, evaluation capacity building, evaluation capacity development, evaluation checklist, evaluation culture, internal evaluation, program evaluation, and nonprofit organizations.

Phases 2 through 4 involved the validation process of the ORECD checklist. Specifically, Phase 2 consisted of a review of the ORECD checklist by relevant experts in order to obtain face and content validity evidences. Phase 3 consisted of pretesting to determine the appropriateness of the wording of the items and format of the ORECD checklist. Phase 4 consisted of a field study in which nonprofit organizations used the ORECD checklist to evaluate their organizations and provided feedback. This phase generated information about the psychometric properties and utility of the ORECD checklist as well as consequential validity evidence. Phases 1 through 4 were conducted separately. Phases 2 through 4 were conducted after the preceding phase was completed, allowing the inclusion of recommendations to the ORECD checklist before conducting the following phase.

Organization of the Study

The dissertation consists of five chapters. Chapter I includes an introduction of the topic, statement of the problem, purpose of the study, research questions, overview of the study, limitations, and definition of terms. Chapter II presents the review of the literature including the framework used to develop the ORECD checklist. Chapter III includes a discussion of the methods that were employed in this study and a detailed description of each of the phases that was conducted. Chapter IV presents the obtained results of the validation process. Chapter V includes the discussion of the findings, recommendations for future research, implications for practice, and conclusions.

Definitions of Terms

Accountability. Is an obligation or willingness by a nonprofit organization to explain its actions to its stakeholders, including government entities, donors, beneficiaries, and the public-

at-large (Dunkle, 2012). This involves deciding both to whom and for what the organizations owe accountability. It is about being held to account by external actors and standards as well as taking internal responsibility for actions (Ebrahim, 2010).

Capacity. The ability of individuals, organizations, and societies as a whole to do their work, solve problems, and set and achieve what is expected in a successful and sustainable manner (Hauge, 1998; OECD, 2006).

Capacity building/development. A process that improves the ability of individuals, groups, organizations, or systems to obtain, strengthen, and maintain the capabilities to set and meet their own objectives or to perform better (LaFond & Brown, 2003; UNDP, 2006). At the organizational level it focuses on developing, supporting, and increasing the organizational infrastructure and required resources to achieve its mission successfully (Taveras et al., 2007). Developing the capacity of organizations or institutions means fostering change within their complex system of policies, enhance and organize their systems, procedures, regulations, resources, and organizational culture and it is seen as mainly endogenous and voluntary driven processes, which means that, it usually focuses on the capacities of organizations, looking from the inside out (Lusthaus, Adrien, & Perstinger, 1999; Simister & Smith, 2010; UNDP, 2006).

Evaluand. The object of evaluation such as program, project, personnel, policy, system, organization, or any other entity being evaluated (Rodríguez-Campos & Rincones-Gómez, 2013; Scriven, 1991).

Evaluation capacity. The ability or potential to conduct evaluation or evaluation activities.

Evaluation capacity development/building. For the purposes of this study, it is defined as “the capacity of putting in place structures that support evaluation efforts within an

organization” (ECDG, 2009a). A thorough discussion of multiples ECD/ECB definitions identified in the literature is presented in Chapter II.

External evaluation. An evaluation conducted by evaluators who are not employees of the organization (Rodríguez-Campos & Rincones-Gómez, 2013).

Formative evaluation. An evaluation typically conducted for the purposes of providing information or feedback for program improvement (Rodríguez-Campos & Rincones-Gómez, 2013; Scriven, 1991).

Internal evaluation. An evaluation conducted by organizational employees or unit from within the organization (Rodríguez-Campos & Rincones-Gómez, 2013; Scriven, 1991).

Internal evaluator. Employee of the organization who performs evaluation functions to any degree, whether alone or in conjunction with other duties and responsibilities (Baron, 2011).

Nonprofit organizations. Also known as not-for-profit organizations and nongovernmental organizations (NGO) are private organizations and separate from government. They exist to contribute to our society, serving a social purpose, a community, or a cause. In nonprofit organizations there are no individual owners who can claim organizational assets for their own benefit. These organizations are not prohibited from creating excess revenue over expenses, but any additional funds must be used to help the organization achieve its goals or mission, not to be distributed as private gain. That means that all profits must be reinvested into the mission of the organization. There are a variety of nonprofit organizations such as: public charities, private foundations, and other exempt organizations such as social welfare organizations, and professional and trade associations (BoardSource, 2010; Carman, Fredericks, & Introcaso, 2008).

Nonprofit organizations 501(c)(3). Nonprofit organizations exempt from federal income tax. They are classified as public charities and private foundations. According to the Internal Revenue Service and the National Center for Charitable Statistics, a public charity is not a private foundation. Public charities receive a substantial portion of their revenue from the general public or the government and include most organizations active in the arts, education, health care, and human services (more than half of all nonprofit organizations fall into this category). Religious organizations are also considered public charities, but they are not required to register with the IRS. A private foundation is often referred to as a non-operating foundation, because typically it does not have active programs. Private foundations are mostly grantmaking or family foundations. The majority supports the work of public charities, provide scholarships, or support government activities.

Organizational capacity. Capabilities, knowledge, and resources that nonprofits need in order to be effective (Connolly & Lukas, 2002). Six components are described as necessary for high performance: governance and leadership; mission, vision, and strategy; program delivery and impact; strategic relationships; resource development; and internal operations and management.

Organizational culture. The norms and values that characterize the operations of the organization (Owen, 2005) and the approaches to management (Volkov & Baron, 2011).

Organizational readiness. The ability and willingness of an organization to adopt and support an enterprise or project.

Organizational readiness for evaluation capacity development. For the purposes of this study, it is defined as the extent to which an organization is prepared with the necessary

components and tools to take on the challenges and embrace the opportunities of developing evaluation capacity.

Policy. A set of rules to guide, regulate and provide direction to organizational actions.

Procedure. Detailed instructions suggesting a sequence of action, including how to and who will implement the policy.

Process. A series of actions, changes, or functions that produce a result or end.

Readiness. State of preparedness of persons, systems, or organizations to meet a situation and carry out a sequence of actions.

Summative evaluation. An evaluation conducted at the end of or after completion of a program for the purposes of providing information to serve decisions or assist in making judgments about program adoption, continuation, or expansion (Rodríguez-Campos & Rincones-Gómez, 2013; Scriven, 1991).

System. A group of interacting, interrelated, or interdependent elements or components forming an integrated whole with a clear purpose (Rodríguez-Campos & Rincones-Gómez, 2013; Weinberg, 2001).

Systematic process. A course of action characterized by order and planning. Usually involves the definition of a problem and searching for possible solutions in order to identify the most appropriate way to address the problem with the expectation that any mistakes or failures will be reduced.

CHAPTER II

LITERATURE REVIEW

This chapter comprises a review of the literature from different sources. The topics presented informed the design and validation of the Organizational Readiness for Evaluation Capacity Development (ORECD) Checklist. The literature review is divided into eleven major parts. The chapter begins by defining the concept of evaluation and explaining the importance of evaluation and evaluation culture. Then, conceptual definitions of evaluation capacity building and evaluation capacity development are presented, followed by a description of the evaluation capacity building/development process, components, models and frameworks, evaluation capacity in nonprofit organizations, internal evaluation, and evaluation checklists.

What is Evaluation?

Evaluation has been defined by different people in different ways. According to Scriven (1991) evaluation refers to “the process of determining the merit, worth, or value of something, or the product of that process” (p. 139). The Joint Committee defined evaluation in its 1994 edition of the *Program Evaluation Standards* as “the systematic investigation of the worth or merit of an object” (p. 3) and in its 2011 edition as “the systematic investigation of the value, importance, or significance of something or someone along defined dimensions” (p 287). Some additional definitions are: “evaluation is a systematic study designed and implemented to determine the value (such as merit or worth) of an evaluand, providing a basis for guiding the decision-making process” (Rodríguez-Campos & Rincones-Gómez, 2013, p. 3); “program

evaluation is a systematic, data-based process for judging the value of a program, helping to make decisions, or creating information about key activities or processes” (King & Volkov, 2005, p. 10); “evaluation is the systematic process of delineating, obtaining, reporting, and applying descriptive and judgmental information about some object’s merit, worth, probity, feasibility, safety, significance, and/or equity” (Stufflebeam & Shinkfield, 2007, p. 16); and evaluation is “the identification, clarification, and application of defensible criteria to determine an evaluation object’s value (worth or merit) in relation to those criteria” (Fitzpatrick, Sanders, & Worthen, 2011, p. 7).

These definitions have some variations among them, however present commonalities as well. According to Stufflebeam and Shinkfield (2007), all of them are based on the meaning of value (evaluation’s root term) to denote that evaluation involves making value judgments. Also, these definitions concur in that evaluation is systematic in determining the merit, value, or worth of an evaluand, including the collection and analysis of information and using defensible criteria to interpret the findings and make judgments (Stufflebeam & Shinkfield, 2007). Specifically, evaluation is presented as a purposeful activity that requires planning and enhances knowledge and decision-making (Preskill & Russ-Eft, 2005; Russ-Eft & Preskill, 2009).

Importance of Evaluation

Evaluation possibly is the most important discipline of society (Stufflebeam & Shinkfield, 2007). It is recognized not only as essential but also as valuable in any system (Fitzpatrick et al., 2011). It is difficult to think of an organization that does not possess an evaluation system (Forss et al., 2006). The main reason to establish an evaluation system is because in the absence of this system there is no information or evidence of high quality that can

be used to improve the program, preventing an organization from accomplishing its mission (Boyle & Lemaire, 1999)

Evaluation has been demonstrated to be a practical tool at the organizational level (Sonnichsen, 1999), up to the point to be integrated into the everyday work of organizations (Russ-Eft & Preskill, 2009). Evaluation is a means to promote transparency and accountability (Malik, 2000) and serves as a management function, a quality assurance mechanism, and as a learning process (Hauge, 1998). It is useful as well to support decision-making providing information about what works, what does not work, and why (Hanwright & Makinson, 2008; Hauge, 1998).

Good quality evaluation is frequently used as a way to help people and organizations (Hauge, 1998) and to achieve better organizational goals in any setting in the public and private sector and at national and international levels (Fitzpatrick et al., 2011; Hauge, 1998). Evaluation is seen now as an ongoing activity that can be used for planning and implementation of a variety of organizational initiatives, instead of an isolated activity that is conducted after a program is over (Hauge, 1998; Russ-Eft & Preskill, 2009).

Evaluation “is a process for giving attestations on such matters as reliability, effectiveness, cost-effectiveness, efficiency, safety, ease of use, and probity” (Stufflebeam & Shinkfield, 2007, p. 4), taking into consideration the day by day realities of organizations and society (Russ-Eft & Preskill, 2009). Therefore, organizations can address questions regarding the performance of a program by using different evaluation processes and tools to obtain valid, reliable, and credible data (Newcomer et al., 2004). Evaluation also serve society by providing affirmations of “improvement, accreditation, accountability, and, when necessary, a basis for terminating bad programs” (Stufflebeam & Shinkfield, 2007, p. 5). It also contributes to

working knowledge and organizational action (Hanwright & Makinson, 2008). According to Russ-Eft and Preskill (2009), “a systematic and professional evaluation adds value to the organization and the work of its members” (p. 10).

The importance of evaluation is not limited to the information that can be obtained to determine merit, worth, or value; it also gives us procedures to improve how we think and, consequently, how we develop, execute, and transform programs and policies (Fitzpatrick et al., 2011). Schwandt (2008) explains that evaluation is useful to determine the effectiveness and efficacy of a program or policy. In this regard Owens (2007) indicates, “evaluation can and should enhance the quality of interventions (policies and programs) designed to solve or ameliorate problems in social and corporate settings” (p. 1). Fitzpatrick et al. (2011) expand on this emphasizing that evaluation can be used not only to improve programs to meet the needs of clients and society but also it is useful to determine how to stick to a budget.

Various authors (e.g., Fitzpatrick et al., 2011; Newcomer et al., 2004; Preskill, 2008; Russ-Eft & Preskill, 2009) also concur in that evaluation is important to facilitate learning. According to Russ-Eft and Preskill (2009), “evaluation is a means for gaining better understanding of what we do and the effects of our actions in the context of society and the work environment” (p. 4). They added that all employees should understand the purpose of evaluation, the processes involved in an evaluation, and its uses. Likewise, Newcomer et al. (2004) point out that evaluation is a valuable learning strategy to increase knowledge in relation to the program theory as well as the practical results of programs. In this regard, Russ-Eft and Preskill (2009) say that in addition to supporting decision-making an evaluation is conducted when some form of learning is needed to take action. They emphasize that an evaluation can

increase knowledge among the members of an organization, which in turn can contribute to building their evaluation capacity.

It is becoming more common to have leadership that wants to lead learning organizations (Newcomer et al., 2004). These authors explain that in learning organizations there are personnel who systematically collect and learn from data about what works and does not work in their program to improve their organizations and the services the organization provides. In other words, evaluation collects data “which are turned into information that, when used, become knowledge at the individual level. If shared with others in the organization, that knowledge may then lead to organization-level learning” (Russ-Eft & Preskill, 2009, p.5). Fitzpatrick et al. (2011) add that evaluators could be helpful to organizations as a whole “through stimulating a learning culture, thereby helping those in the organization to question and consider their goals and their methods, their clients and their needs, and showing them how to use evaluative inquiry methods to meet their needs” (p. 33).

Evaluation Culture

Evaluation culture is described by Owen (2003) “as a commitment to roles for evaluation in decision-making within an organization” (p. 43). When there is a lack of commitment, both evaluation and the development of evaluation capacity could be overshadowed by other activities (Baron, 2011). Thus, a culture that supports evaluation activities has been identified as an important precursor of ECB (Boyle & Lemaire, 1999; Owen, 2003). Russon and Russon (2007) propose three dimensions as indicators of a culture of evaluation in an organization. These dimensions include what they called artifacts (e.g., records, reports, and products); values or beliefs about what is appropriate or not, laying the foundation for the artifacts; and assumptions or beliefs about the evaluation that provide the foundation for the values (Russon & Russon,

2007). They explain that values and assumptions are the result of personal and other people experiences.

Furthermore, ECD needs to be directed to the organization and not just the personnel who are likely to undertake evaluations, in order to develop and sustain an evaluation culture (Beere, 2005). Baron (2011) indicates that, “Whether an organization is an office of 2, 10, or 50, evaluation needs to be discussed actively if it is to be prevalent throughout the organization’s program” (p. 90). Understanding the evaluation culture of an organization is essential for understanding the development of evaluation capacity (GAO, 2003; Baizerman, Compton, & Stockdill, 2002b).

Evaluation culture includes formal, regular processes in place, to plan, execute, and use information from evaluations (GAO, 2003). Its development includes participation, dialogue, and skill building development, among others (Kirsh, Krupa, Horgan, Kelly, & Carr, 2005). “If the organization already has a culture where members freely share information, trust one another, consistently ask questions, and take risks, then it is more likely that ECB efforts will be successful” (Preskill & Boyle, 2008, p. 453). The establishment of an evaluation culture in nonprofit organizations is a gradual process in which members identify how to use information from evaluations (Bozzo, 2002). It requires an examination of the characteristics of the organization, acceptance by personnel, and organizational work practices (Hanwright & Makinson, 2008).

The development of internal evaluation capacity is important to foster an evaluation culture (Dabelstein, 2003). According to Owen (2003), internal evaluation and evaluation culture are related in the following ways:

An internal evaluation regime is consistent with an organization becoming a center of inquiry. In such a regime, we no longer think of organizations simply as knowledge distribution centers. An organization must be concerned with more than delivery; it must also be a producer as well as a transmitter of knowledge. One can think of an organization with this perspective as engaged in a pervasive search for meaning in its work. If this position is adopted, then the organization has developed a culture of evaluation. (p. 44)

The shift toward more flexible, internal evaluations “requires that organizations develop a culture that understands and appreciates the value of evaluation in order to help ensure evaluation success” (Arnold, 2006, p. 258). It also calls for people to see evaluation as an integral part of the organization rather than an isolated activity (Russon & Russon, 2007). Owen (2007) indicates that two key factors affect the development of an evaluation culture: roles of management, including support from the top, and effective change and innovation strategies. Hanwright and Makinson (2008) concur in that the development of evaluation requires organizational change. Milstein et al. (2002) explain that a process of culture change is required to strengthen evaluation capacity and suggest a variety of opportunities for culture change. Some of the most relevant are: good understanding of the meanings of evaluation and the use of a common language, clear understanding of program theory using logic models or other graphical representations, use of diverse methods of data collection, modeling of good evaluation practice, and enhancing internal capacity for evaluation through new directions in leadership, funding, training, and information sharing, among other opportunities (Milstein et al., 2002).

Evaluation Capacity Building/Development Definitions

The concept of ECB/ECD is broad and different from specific roles, evaluation approaches and program evaluation per se (Baizerman et al., 2002b; King, 2007). ECB is a relatively recent conceptual development, that is just emerging, intended to tackle the program evaluation fundamental problems (King & Volkov, 2005; Nielsen, Lemire, & Skov, 2011; Trevisan, 2002). Program evaluation usually focus in the completion of a single evaluation study and the improvement of a program; and not in the development of the infrastructure and processes necessary for sustaining evaluation and improving the organization that sponsor the program, in order to institutionalize the evaluation practice (Compton, 2009; Stockdill et al., 2002).

ECB has been recognized as “a multidimensional construct involving maximizing potential in different areas” (Kirsh et al., 2005, p. 235). Beere (2005) indicates that ECB involves not only the expertise needed to conduct high-quality evaluations but also promotes an organizational culture in which evaluation is part of the everyday work. King and Volkov (2005) are more specific saying:

The goal of ECB is to strengthen and sustain effective program evaluation practices by increasing an organization’s capacity to: design, implement, and manage effective evaluation projects; access, build, and use evaluative knowledge and skills; cultivate a spirit of continuous organizational learning, improvement, and accountability; and create awareness and support for program evaluation and self-evaluation as a performance improvement strategy in the internal and external environments in which they function.

(p. 11)

Various definitions of ECB and ECD have emerged based on different research contexts resulting in varying conceptualizations; however, they share some common characteristics or traits (Labin, et al., 2012; Naccarella et al., 2007; Tseng, 2011). For example, some definitions emphasize the notion of equipping organizations to conduct ECB, evaluations, and others processes and address the role of ECB facilitating the use of evaluation (Levin-Rozalis et al., 2009; Naccarella et al., 2007). Also, there are definitions that focus on building evaluation capacity at the individual and organizational levels or only at the organizational level (Labin, et al., 2012).

Compton and Baizerman (2007) indicate that the diversity of ECB definitions is something expected and encouraged to promote the development of new ideas and explain that “Early efforts at control, at defining boundaries and asserting orthodoxy, may serve to restrict, deflect, or push underground the vibrant exploration of idea and practice” (p. 118). According to Taut (2007), it is complex to define an idea like ECB because evaluation capacity is highly context dependent and its conceptualization is diverse; therefore, a definition that is applicable or useful in one setting perhaps is not appropriate in another setting. A variety of ECD and ECB definitions that can be found in the literature are as follows.

Evaluation Capacity Development

According to Boyle and Lemaire (1999) evaluation capacity refers to “the human capital (skills, knowledge, experience, etc.) and financial/material resources” (p. 5) and evaluation capacity development refers to “activities and initiatives taken to implement an evaluation regime” (p. 6). The authors describe evaluation regime as “the configuration of evaluation capacity, evaluation practice, organizational arrangements, and institutionalization” (p. 6).

Another ECD definition includes the one offered by Mackay (1999) in which ECD is defined as

“the development of national or sectoral evaluation systems” (p. 2). According to Picciotto (1998), ECD is “the ability of public institutions to manage information, assess program performance, and respond flexibly to new demands” (p. 39).

ECD is also defined by Khan (1998) as “an activity or a set of activities that contribute to the establishment of evaluation capacities within the development administration structures of developing countries” (p. 312). Khan (2000) also explains that the evolution of systems and methodologies is essential for the development of evaluation capacities to “assist lessons learning from on-going or past projects and programs and, through these lessons, adjust projects and programs in such a manner that they achieve their planned objectives or improve the quality of design of similar projects in the future” (p. 11). The most recent definition of evaluation capacity development is the one provided by the Evaluation Capacity Development Group (2009a) in which ECD is defined as “the capacity of putting in place structures which support evaluation efforts within an organization. This process evolves into the creation of an evaluation system” (“What is Evaluation Capacity Development?” para. 1).

Evaluation Capacity Building

Among the most prevalent definitions of ECB is the one by Stockdill, Baizerman, and Compton. ECB is conceptually defined by these authors as a “context-dependent, intentional action system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/programs/sites” (2002, p. 8). These authors also developed a working definition in which ECB is defined as “the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and

its uses routine” (2002, p. 14). Another broad description of the ECB concept is presented by Schaumburg-Müller (1996):

It includes activities, which provide support for systems of evaluation, audit, feedback, and learning from policies, programs or projects performed at various levels, mainly in the public sector. Although the concept is defined broadly, it excludes activities aimed solely at planning and appraisal activities. Also, the interest focuses on activities which are not just of a temporary nature but have the aim of supporting a sustainable evaluation function. Therefore, support for temporary monitoring and evaluation units connected with a specific aid activity is excluded unless it provides evaluation training of a more general and sustainable nature to host-country staff. (p. 5)

Gibbs et al. (2002) define ECB as the “the extent to which a community-based organization has the necessary resources and motivation to conduct, analyze, and use evaluation” (p. 261). The definition by Bozzo (2002) describes ECB “as the development of resources, technical skills, and understanding to enable organizations to undertake evaluation activities” (p. 77) and Beere (2005) defined it as “the ability to conduct an effective evaluation” (p. 41). More recently, Preskill and Boyle (2008) presented a comprehensive definition saying that:

Evaluation capacity building involves the design and implementation of teaching and learning strategies to help individuals, groups, and organizations, learn about what constitutes effective, useful, and professional evaluation practice. The ultimate goal of ECB is sustainable evaluation practice – where members continuously ask questions that matter, collect, analyze and interpret data, and use evaluation findings for decision-making and action. For evaluation practice to be sustained, participants must be provided with leadership support, incentives, resources, and opportunities to transfer their learning

about evaluation to their everyday work. Sustainable evaluation practice also requires the development of systems, processes, policies, and plans that help embed evaluation work into the way the organization accomplishes its mission and strategic goals. (p. 444)

In 2012, another working definition of ECB emerged. Labin et al. (2012) defined it as “intentional process to increase individual motivation, knowledge, and skills, and to enhance a group of organization’s ability to conduct or use evaluation.” (p. 308).

Evaluation Capacity Building/Development Process

ECB as an emergent area of practice has experienced an increased interest from governments and organizations that seek to enhance their effectiveness and accountability (Baizerman, Compton, & Stockdill, 2002a; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005). For example, Compton, Glover-Kudon, Smith, and Avery (2002) indicate that this increased interest in evaluation from stakeholders supported the ECB efforts within the American Cancer Society to co-create and co-sustain evaluation as a standard process and everyday practice. Nevertheless, ECB is difficult to achieve and has been described as “a complex phenomenon involving issues of individual learning, organizational change, sustained change, and program processes and outcomes” (Labin et al., 2012, p. 328). As ECB matures more research and attention is needed to better comprehend what it takes to successfully sustain this process, identify the learning theories behind it, gain understanding of its practices, and measure its impact (Huffman et al., 2008; Stockdill et al., 2002; Suárez-Balcázar & Taylor-Ritzler, 2014; Taylor-Powell & Boyd, 2008; Wandersman, 2014).

Compton (2009) indicates, “it is too early in the history of ECB to have empirically based principles of practice” (p. 66). Specifically, Baizerman et al. (2002) explain that the ECB process is not completely detailed and includes somewhat unclear practices and little systematic

assessment. One of the concerns according to Milstein et al. (2002) is that “the evaluation profession as a whole still lacks a well-developed theory and associated indicators for understanding evaluation capacity at an organizational level, particularly its inherent change over time and ‘ongoingness’” (p. 41). Therefore, it is important to present models for building evaluation capacity and testing them in real settings in order to strengthen and move the concept forward (Huffman, Lawrenz, Thomas, & Clarkson, 2006).

Many authors (Compton, 2009; Compton et al., 2002; Gilliam et al., 2003; Huffman et al., 2008; LaFond & Brown, 2003; Mackay, 2002; Milstein et al., 2002; Sridharan & De Silva, 2010; Stockdill et al., 2002) concur in that ECB is more than a single process or practice; it is a family of ongoing, dynamic, evolutionary, and ever changing guided processes, multiple practices, and multiple strategies to develop and sustain quality program evaluation and its appropriate uses, and must proceed through activities that are self-perpetuating. These evolutionary processes that allow the development of evaluation capacity into an evaluation system have been described as non-linear, meaning that the elements of the system have different types of relationships between each other (ECDG, 2009a). King and Volkov (2005) indicate that this system of guided processes and practices “necessarily includes a wide variety of adult learning processes, requiring active participation of learners during the entire evaluation process” (p.15).

ECB is about the development of sustainable evaluation efforts that can be located at different levels (Stockdill et al., 2002). At the organizational level, many things need to be implemented (Valéry & Shakir, 2005). The development of evaluation capacity in organizations is a complex and multifaceted task (Arnold, 2006); it is challenging and includes a broad range of difficulties and issues (Huffman et al., 2008). Because of the complexity of the ECB practice,

“the ongoing structures and processes of managing ECB require an unusually high degree of continuous and empirical monitoring to ensure that the complex action system continues to be focused, appropriate, and effective” (Compton, 2009, p. 66). Even though developing evaluation capacity is complex, King and Volkov (2005) recognize that it “can be built, slowly and systematically over time, through procedures that make sense even for small organizations” (p. 12).

ECB should be an organizational practice and a priority to the organization’s work (Compton et al., 2002; Forss et al., 2006). This means that ECB is not just about developing the skills and knowledge of individuals; it goes further focusing on working with the whole organization, developing the skills appropriate for each level of an organization, and building awareness of techniques and approaches that are practical and feasible (King & Volkov, 2005). Besides, the ECB processes emphasize the development and sustainability of a long-term infrastructure as a way to support short-term evaluation studies (Compton, 2009). As this author explains, “to do ECB someone must focus beyond the immediate study, beyond the process, completion, and timely use of a single evaluation study toward the multiple activities necessary for creating demand for evaluation as a regular and routine part of the organization’s work” (p. 66).

In order to establish or strengthen an internal evaluation system through developing evaluation capacity it is necessary first, to conduct a thorough analysis and examination to identify the existence and the status of that system in the organization and its future sustainability (Khan, 1998). García-Iriarte et al. (2011) explain that it is necessary to assess the organization’s readiness before engaging in the ECB process. They suggest as part of the assessment, a review of organizational practices, processes, evaluation reports and other documents as well as gaining

an understanding of the staff capability and their motivation and expectations for the evaluation. Baizerman et al. (2002a) indicate that the assessment may also include the organizational wants, needs, and dreams for evaluation. Hauge (1998) recommends, when possible, “to build on existing systems and institutional realities and to use and strengthen existing capacities, rather than starting from scratch” (p. 31) because that increases the likelihood of success.

Furthermore, there is a strong need to develop a receptive culture that supports the development of evaluation capacity, in which demand for evaluation, effective use of evaluation outputs, and ownership of evaluation findings can grow (Forss et al., 2006; King & Volkov, 2005), because this will provide lasting organizational benefits (McDonald et al., 2003). According to Stockdill et al. (2002) ECB is clearly oriented to structures, cultures, and everyday practices that occur within any individual organization, between programs in the same organization, or between similar organizations. In this regard, Preskill and Boyle (2008) propose a set of ECB assumptions that organizational leaders and evaluators might make and share with other key leaders to emphasize that an ECB effort can be successful and valuable. These assumptions are: “(a) organization members can learn how to design and conduct evaluations, (b) making learning intentional enhances learning from and about evaluation, and (c) if organization members think evaluatively, their programs will be more effective” (p. 446). These assumptions take for granted that evaluation is something good to do, add value to the organization, and can contribute to an effective decision-making process (Preskill & Boyle, 2008).

Evaluation Capacity Building/Development Components

The success of the ECB process depends on organizational commitment in different areas (Adams & Dickinson, 2010). This process “involves the creation and maintenance of an evaluation environment within the organization” (Baron, 2011, p. 88). Key components known

to contribute to developing evaluation capacity within organizations according to the literature are as follows.

Organizational Environment

How ECD practices are conducted is highly dependent on the context of each particular organization, the organizational culture and history, everyday ways of working, organizational mission, type of organizational structure, and how evaluation and evaluators are perceived (Baizerman et al., 2002a; Compton, 2009; King, 2007; King & Volkov, 2005; Love, 1983; O’Sullivan & O’Sullivan, 1998; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005). It is necessary to tailor the ECB efforts according to the circumstances of the organization; therefore a “cookie-cutter” ECB approach would be ineffective (Mackay, 2002). Stockdill et al. (2002) concur with Mackay, indicating that because “ECB is profoundly contextual in terms of structure, culture, resources, politics, and ideologies of the organization, program, or other site, any effort to bring in a ‘canned’ ECB structure or process is likely to fall, particularly in the long run” (p. 21).

Having an understanding of the internal and external organizational contexts, the organizational culture, and everyday ways of working is necessary to determine whether building evaluation capacity is feasible, develop methods to build evaluation capacity, and identify strategies (e.g., commitment to learning from evaluation, specialized data systems and analysis, and creative ways to get funding) to support the development of evaluation capacity to make it sustainable (Arnold, 2006; GAO, 2003; King, 2007; King & Volkov, 2005; Newcomer, 2004). As Milstein and colleagues (2002) state, “the movement to build evaluation capacity for evaluation could not proceed as an isolated initiative” (p. 36). Becoming involved in the organizational culture allows the development of evaluation capacity (Baron, 2011; Russon &

Russon, 2007). It is also essential to have the presence of evaluation champions who are committed to evaluation because these can be influential, particularly if these champions are part of the organizational leadership (Cousins & Bourgeois, 2014).

King (2007) and King and Volkov (2005) point out that it is important to understand the evaluation pressures from the external environment (e.g., legislation's requirements, funders' requirements) and the extent to which there is potential support for change. They also added that it is essential to perform a preliminary assessment of the internal environment to determine its readiness, including the existing organizational learning capacity, existing internal evaluation capacity, support for previous evaluation work, and organizational stability. The lack of organizational stability is one of the major barriers to build evaluation capacity (Stevenson et al., 2002) and can be detrimental to the organizational commitment to participate in ECB activities (Atkinson et al., 2005). According to King and Volkov (2005), certain characteristics of an encouraging environment to support ECB include open mindedness, lack of fear (of being penalized), respect for each other, rewards for innovation/risk-taking/creativity, a sense of humor, and positive attitudes toward evaluation.

Organizational Leadership Support

Organizational leadership support or administrative support is defined by Majchrzak (1982) as a "commitment of the administration to the use of evaluation information in managerial decision making and planning" (p. 308). This author explains that evaluation tends to be integrated at the decision making level when the leadership provides a high degree of support to the organization. Many authors (Adams & Dickinson, 2010; Compton, 2009; Cousins & Bourgeois, 2014; Dabelstein, 2003; Kapucu, Augustin, & Krause, 2007; King, 2002; Russon & Russon, 2007; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008) concur in that engagement

and support from the organizational leadership is critical to continued development, sustain ECB, and to integrate evaluation into organizational life.

The leadership is in charge of making decisions regarding staff, budget, promotions, and identify what they want to be evaluated (Compton, 2009). In the absence of active support, the everyday demands of required studies and continuous activities eliminate the possibility of capacity building (King & Volkov, 2005; Taut, 2007b). Also, the lack of leadership is a major barrier to developing evaluation capacity (Taut, 2007b) and impedes learning from the ECB process (Forss et al., 2006). In contrast, García-Iriarte et al. (2011) found that staff become more engaged in evaluation practices and develop more ownership as a result of leadership support. Thus, a long-term commitment from the leadership is necessary to develop evaluation capacity (Khan, 1998).

At least, administrators should agree not to interfere with evaluation activities; however, it is better when they support the ECB process as role models who evaluate their own activities, express their support to others, and encourage involvement (King, 2007; King & Volkov, 2005; Taylor-Powell & Boyd, 2008). Equally important is to identify in an organization “how evaluation training and/or planning is valued by upper levels” (Newcomer, 2004, p. 214). Also, the organizational leadership needs to believe those efforts can provide positive rewards otherwise they may not support the effort (Newcomer, 2004). In this respect Hoole and Patterson (2008) indicate:

Capacity building efforts by funders are important, but commitment of organizational leadership is critical in transforming the role of evaluation from one of basic reporting and accountability to a true process of continuous organizational learning. This

transformation requires leadership commitment to development of an evaluative learning culture and development of an infrastructure to support it (p.111).

Knowledge/Skills Development

When building evaluation capacity, the aim of an organization should focus on the need to develop some level of evaluation expertise within the organization (Sonnichsen, 1999; Preskill, 2014). The development at the individual and organizational level is essential (Atkinson, Wilson, & Avula, 2005). All organizational levels, not only the top leadership or a specialized evaluation unit, must have access to formal training and professional development in evaluation, mentoring and coaching, technical assistance, written information, meetings, technology, and involvement in interactive evaluation processes, in order to promote the value and understanding about evaluation, engage people in the ECB process, and support the organization continued effort to obtain data for decision-making (Adams & Dickinson, 2010; Atkinson et al., 2005; Dabelstein, 2003; Duignan, 2003; Forss et al., 2006; King, 2002; King & Volkov, 2005; Monroe et al., 2005; Preskill, 2008; Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008).

Knowledgeable people trained in evaluative thinking and skills are more likely to improve and gradually increase their participation in the ECB process, increase their capacity to manage the ECB process as well as to conduct and use evaluation, make more timely and effective decisions, and adapt to changing conditions more effectively (Forss et al., 2006; O'Sullivan & O'Sullivan, 1998; Preskill & Boyle, 2008; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005). In order to be successful, the development of knowledge and skills should employ strategies that engage people in collaborative learning and experiences in addition to the traditional formal presentations (Huffman et al., 2008; Preskill, 2008). Mutual

learning, knowledge transfer to everyday work, and improvement of attitudes and beliefs about evaluation are essential to building evaluation capacity (Kirsh et al., 2005; Preskill, 2014; Preskill & Boyle, 2008). Also, Arnold (2006) highlights the need to have in-house evaluation expertise to build evaluation capacity.

Employee turnover is another reason why the organization must identify and provide ongoing opportunities to those who are interested in evaluation, in order to grow the pool of new supporters on a regular basis, sustaining the continuation of evaluation in the organization, because making evaluation a regular and consistent practice is difficult when turnover is high (Bozzo, 2002; Compton et al., 2002; King, 2007; Labin et al., 2012; Preskill & Boyle, 2008; Taylor-Powell & Boyd, 2008). Gibbs et al. (2002) went beyond saying that training should include as many staff members as possible (not just the interested people) to support the continuity and advancement of the evaluation efforts. In order to build an evaluation-literate staff and grow the pool of potential evaluators, it is important to understand in advance, whose capacity should be developed, the level of evaluation knowledge and skills of potential participants, and their beliefs and motivation toward evaluation (Forss et al., 2006; Milstein et al., 2002; Preskill & Boyle, 2008). In addition, support and commitment from the organizational leadership is crucial (Stevenson et al., 2002).

Although, ECD typically include professional training and instruction for individuals, "...it goes further by attempting to extend the development of individuals to affect the future work of the organization" (Huffman et al., 2008, p. 359). Specifically, these authors explain that the growth of evaluation capacity of individuals could increase the capacity of the organization as a whole, which at the same time will make the organization able to provide more support for the individuals to expand and grow in evaluation capacity.

Resources

In order to develop evaluation capacity effectively it is important to determine the available resources of the organization in order to assure their adequacy (Adams & Dickinson, 2010; Gibbs et al., 2002; Labin, 2014; Milstein et al., 2002; Preskill & Boyle, 2008). Funding for evaluation activities, personnel, time to collaborate on evaluation activities (particularly during the workday), tools and technology such as foundation for data collection and analysis, materials, equipment, space, computer hardware, software, printers, and databases are among the resources necessary to sustain evaluation capacity development practices (Dabelstein, 2003; Gibbs et al., 2002; King, 2002; King & Volkov, 2005; Preskill & Boyle, 2008; Russon & Russon, 2007).

Numerous authors (García-Iriarte et al., 2011; Kirsh et al., 2005; Lennie, 2005; O’Sullivan & O’Sullivan, 1998; Preskill & Boyle, 2008; Stockdill et al., 2002) have identified lack of organizational infrastructure, that is, lack of human and financial resources, appropriate technologies, and time as the main barriers that place major constraints to developing and implementing evaluation capacity. Specifically, financial and human resources are considered essential to support ECB and their absent is considered a serious barrier (Arnold, 2006; King & Volkov, 2005; Kirsh, 2005). When there is no allocation of specific resources or a financial commitment in place early in the process to develop evaluation capacity, pressure is frequently present because of the cost of evaluation (Russon & Russon, 2007) and “tension may develop between spending time and resources toward data collection and producing high-quality evaluation reports, and implementing activities and processes designed to help participants learn from and about evaluation” (Preskill & Boyle, 2008, pp. 448-449).

Program Theory

One way to build evaluation capacity within organizations and assist personnel to think about their program(s) and the outcomes they can achieve is by articulating the program theory and logic models (Lambur, 2008; Monroe et al., 2005). These are useful to “be able to get what evaluation and performance mean to organizational leaders...” (Volkov & Baron, 2011, p. 102). Organizational leadership, personnel, and stakeholders need to understand the fundamentals of program theory and the evaluation process before initiating it (Monroe et al., 2005). This is essential to ensure the quality of data collected, including credible and reliable information to support conclusions on program effectiveness (GAO, 2003).

Program theory as defined by Bickman (1987) is “the construction of a plausible and sensible model of how a program is supposed to work” (p. 5). The model should be well-defined to avoid the risk of implementing an evaluation that does not meet the needs of an organization (Adams & Dickinson, 2010; Gugiu & Rodríguez-Campos, 2007). Specifically, a logic model is useful to illustrate a program’s theory or make it explicit, exemplify how a program works, help an organization specify intended levels of achievement, and it is a means to make ECB assumptions, expectations, and roles clear (Carman, 2007; Compton et al., 2002; Connolly & York, 2002; Frechtling, 2007; Taylor-Powell & Boyd, 2008).

Logic models are graphic displays that exemplify how to move from the current conditions to the vision for success (Suárez-Balcázar & Harper, 2003). Thus, the construction of a logic model is recommended as an important initial step before the implementation of an evaluation (Gugiu & Rodríguez-Campos, 2007), in order to have a better understanding of the underlying principle behind the program’s intended effects (Fitzpatrick et al., 2011). Logic models lay the foundation for a broad and meaningful evaluation (Frechtling, 2007).

The use of logic models are endorsed and required by many organizations that require evaluation, including government agencies, the United Way, and prominent foundations (Carman, 2007; Fitzpatrick et al., 2011; Kaplan & Garrett, 2005), as a tool to define and make clear what needs to be measured and when, describing the sequence of the events (Frechtling, 2007). The utility of logic models is highlighted by King (2002) in a study conducted within a school district, showing that the difficulties to understand the link between a strategy and its outcome were associated to the lack of logic models.

Demand for Evaluation

Demand for information from external stakeholders on what works, accountability requirements, and/or internal needs such as a desire to increase evaluation use and funding sources are usually the driving force and main prerequisite to develop a commitment for building evaluation capacity (GAO, 2003; Mackay, 2002; Preskill & Boyle, 2008). The demand for evaluation as part of ECD involves a conscious effort, meaning that it requires to be created and sustained (Mackay, 2002; Rotondo, 2012) and its sustainability “depends on the extent to which evaluation is used within the organization” (Cousins, Goh, Elliott, & Bourgeois, 2014, p. 10). In other words, demand is the result of practical need (Hauge, 1998; Stockdill et al., 2002) and emerges when people want evidence of results for some rational purpose (Hague, 1998; Taylor-Powell & Boyd, 2008). It is important to make sure that there is demand to develop evaluation capacity and a plan about how to accomplish it before embarking in this process (Forss et al., 2006). Without the demand or understanding of the value of evaluation or what evaluation represents and the need to conduct evaluation activities, there is little point to developing evaluation capacity (Hauge, 1998; Satterlund et al., 2013). Thus, the frequent lack of genuine

demand for evaluation, together with lack of evaluation, accounting or auditing skills are among the main barriers for developing evaluation capacity (Hauge, 1998; Mackay, 1999).

Communication

The existence of an effective organizational communication system is critical in increasing the probabilities of collaboration in the ECB process (King & Volkov, 2005). The success of developing evaluation capacity “depends on communication structures that facilitate horizontal and vertical information flow across the entire organization” (Taylor-Powell & Boyd, 2008, p. 63). Lack of transparent communication or no communication at all, is an organizational factor that impedes learning from the ECB process (Forss et al., 2006; King, 2002).

Policies and Procedures

An essential factor to institutionalized evaluation in an organization is the development of policies and procedures, which provide the necessary structures and means to embed evaluation in a way that allows the organization to achieve its mission and goals (Adams & Dickinson, 2010; Preskill & Boyle, 2008; Russon & Russon, 2007). These policies and procedures could consist, for example, of explicit rules and procedures to guide evaluation decisions and actions and how evaluation will be used, requirements and incentives to engage in evaluation, expectations about routine activities, and explicit evaluative roles (King, 2005; Russon & Russon, 2007; Taut, 2007b; Taylor-Powell & Boyd, 2008).

An explicit evaluation capacity building policy, as part of good governance initiatives, should be established in the organization to avoid any implicit process, such as personal decisions, and clearly communicate what the expectations are (Dabelstein, 2003; Taut, 2007b;

Valéry & Shakir, 2005). This might include the establishment of an evaluation unit or team responsible for the ECB process, in addition to the development of internal processes to make evaluation part of the daily work practices of the organization (Preskill & Boyle, 2008; Russon & Russon, 2007).

Some of the elements that are typically part of the policies in an organization are the ones regarding evaluation training for staff, disclosure of evaluation information, and ethical considerations (Duignan, 2003; Russon & Russon, 2007). Some of the most common evaluation procedures include sources to obtain technical assistance in evaluation, guidelines on the use of internal and external evaluators, and stakeholder consultation standards for evaluation planning (Duignan, 2003; Russon & Russon, 2007).

External Support

The development of evaluation capacity does not mean that the organization is in charge of all the evaluation work (Taylor-Powell & Boyd, 2008). Developing internal evaluation capacity will be equally important to ensure that when external resources are solicited they are a good match for the needs of the organization (Newcomer, 2004). External support and collaboration also imply a more proactive participatory role in capacity building from private and public agencies, organizations, and evaluators (Bozzo, 2002).

In order to support the ECD process organizations may have access to personnel in the form of internal professionals, establish collaborative relationships with external experts or partners, engage partners, build networks, and participate in communities of practice (King & Volkov, 2005; Preskill & Boyle, 2008; Russon & Russon, 2007; Satterlund et al., 2013; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005). These are opportunities to develop evaluation

capacity by sharing knowledge, skills, experiences, information, goals, and beliefs about evaluation (GAO, 2003; Monroe et al., 2005; Preskill, 2008; Preskill & Boyle, 2008).

Peer support can build on existing mechanisms to facilitate ECB as well (Taylor-Powell & Boyd, 2008). Specifically, benchmarking, sharing evaluation plans and findings, learning from other organizations and/or working together with them can be ways for advancing evaluation, promote the exchange of ideas, and support self-assessment when organizations compare their ideas and strategies to the ones of other organizations, which could influence an organization to make major positive changes (Bozzo, 2002; Carman, 2007; GAO, 2003; Mott, 2003).

Incentives

Incentives are identified by Adams and Dickinson (2010) as one of the key factors known to contribute to developing evaluation capacity. Boyle and Lemaire (1999) suggest that it is essential to know what incentives are available to encourage the development of evaluation capacity. ECB should include meaningful incentives for participation in evaluation such as the time and flexibility that people need to incorporate the evaluation process into their daily work and personnel acknowledgement by the organization (King & Volkov, 2005). This will serve as a motivator for participants as well as emphasizing to learners that participation in the evaluation process can lead to the development of valuable, lifelong skills (King & Volkov, 2005). The motivation to be involved in an evaluation facilitates the learning process (Forss et al., 2006). Incentives and motivation are important because the staff in an organization already have many responsibilities, so they need to know what evaluation can offer them if they are going to embark in something else (King, 2002).

Feedback Mechanism

Feedback is an important component of developing evaluation capacity; however, it usually gets relatively little attention (Khan, 1998). Lack of a feedback mechanism on the effectiveness of the organization and the decision-making processes is one of the major barriers to develop evaluation capacity (Hauge, 1998; Mackay, 1999). Feedback facilitates the learning process within the organization and between organizations (Bozzo, 2002). It is also the link between the use of evaluation and the decision-making process (Khan, 1998). A feedback mechanism in the decision-making process is essential to increase the possibility of collaboration in the ECB process (King & Volkov, 2005). Similarly, when leaders are receptive to feedback from others, there is an increased likelihood to have a more lasting impact from any ECB process and future evaluation efforts (Preskill & Boyle, 2008).

Evaluation Use

The use of evaluation has been recognized as an important aspect and outcome of ECB (Clinton, 2014; Cousins, Goh, Elliott, & Bourgeois, 2014; Dabelstein, 2003; Preskill & Boyle, 2008). According to Naccarella et al. (2007), it is necessary for ECB to “stress the varied uses to which evaluation findings should be put” (p. 235). When there is no utilization “the ECB efforts will have failed” (Mackay, 2002, p. 90). The use of evaluation findings could be achieved if evaluation supports activities through learning which, ultimately, support the sustainability of evaluation capacity (Bozzo, 2002; Mackay, 2002). Learning from evaluation allows continuous improvements of the organization (GAO, 2003), and it occurs through interaction with others, including practice and active involvement (Taut, 2007b).

Bozzo (2002) and Hauge (1998) indicate that findings should be understandable and usable by organizations. They explain that end-users need to feel comfortable with the form the

data is presented to enhance the utilization of the findings and act upon them to fulfill their function. Similarly, Patton (2008) says that when people are actively involved in evaluation it is easier for them to understand and feel ownership of the evaluation process which increases the possibility to use evaluations.

People in the organization must be able to use evaluation results to make appropriate decisions and changes (King, 2002; King & Volkov, 2005). It is important not only to use evaluation findings for decision-making but also to communicate the uses of these findings (Preskill & Boyle, 2008). Therefore, it is helpful to know early on, the extent to which people who will participate in ECD activities have sufficient input or power into organizational decision-making and whether they are unable to use evaluative information due to existing power structures in the organization (King, 2007).

Evaluation Capacity Building/Development Models and Frameworks

Various ECB/ECD models, approaches, and frameworks have been developed in many contexts. The differences between them are related to “depth and breadth of the ECB initiative, ECB connected to ‘conducting and evaluation’ versus broader evaluative inquiry, and intentional versus more opportunistic approaches to ECB” (Taylor-Powell & Boyd, 2008, p. 66). In addition to the ECB/ECD models, approaches, and frameworks there are evaluation approaches that support the development of evaluation capacity such as participatory, collaborative, and empowerment evaluation; however, these are not the focus of this chapter discussion (for more information about these approaches, see for example, Adams & Dickinson, 2010; Atkinson et al., 2005; Compton, Baizerman, Preskill, Rieker, & Miner, 2001; Díaz-Puente, Yagüe, & Alfonso, 2008; Huffman et al., 2006; Khan, 2002; King, 2007; Kuzmin, 2012; Lennie, 2005; Milstein et al., 2002; Rodríguez-Campos & Rincones-Gómez, 2013; Taut, 2007b; Trevisan, 2002).

According to Nielsen et al. (2011), some models, approaches, and frameworks focus more on the development of human capital, tools, and resources, and others emphasize the demand side, such as organizational structures, policies, and processes. It is important to note that ECD can be addressed at the individual, organizational, and societal level (Zhaoying, 2003). However, the following models, approaches, and frameworks identified in the literature and presented in this chapter mainly focus on developing evaluation capacity at the organizational level. There are some cases in which elements to develop evaluation capacity at the individual level are present.

Evaluation Capacity Development Group Toolkit

The second edition of the *Evaluation Capacity Development Group Toolkit* [ECDG Toolkit] developed by Russon and Russon (2007), “focuses on how organizations can include evaluation in the design of their jobs in order to better meet important individual needs while contributing to increased organizational effectiveness” (ECDG, 2009c, “Integrating Evaluation into Individual Jobs,” para. 2). The Toolkit contains 10 tools. The first six tools are about how to develop evaluation capacity and the last four tools consist of the application of what was learned in the first six tools, in addition to other resources. The 10 tools are: (a) to rent or to own – to determine when is appropriate for an organization to hire an external or an internal evaluator; (b) purpose and shared vision – to determine the purpose of the evaluation and create a shared vision for evaluation that reflect the desired future; (c) organizational design – to put in place a design that enables the organization to achieve the shared vision; (d) organizational culture – to help incorporate evaluation into the organizational culture to promote the development of a culture of evaluation; (e) evaluation policies – to determine if the existing or proposed policies are having the desired results; (f) budget – to estimate the evaluation budget

and estimate the correct amount; (g) processes – to start using the newly developed evaluation capacity; (h) approaches to training – to assess the need for training, determine the instructional goals and method, and training implementation and follow up; (i) action research – to identify if the planned changes in the organization are working or not; and (j) standards for internal evaluation – to determine how well the organization is using its new evaluation capacity (Russon & Russon, 2007).

Grounded Framework for Evaluation Capacity Building

This framework developed by King and Volkov (2005) is based on lessons learned from a study with three organizations. Its objective is to support and institutionalize evaluation in organizations. According to the authors, it was originally designed for nonprofit organizations seeking to increase their long-term capacity to conduct and use program evaluations in everyday activities; however, it can be useful for other organizations that want to improve the quality and quantity of their evaluations. The framework “provides a common and consistent approach to developing a practical evaluation function when planning and implementing organizational performance improvement strategies and accountability mechanism” (King & Volkov, p. 12). It consists of three major categories and each category consists of several components. The categories are: (a) organizational context – consists of the external organizational context which locates an organization in time and place and the internal organizational context which is key to determine the feasibility of ECB; (b) ECB structures – development of these structures will create mechanism to build evaluation capacity; and (c) resources – organizations must have easy access to evaluation resources, and the sources of support for program evaluation in the organization should be explicit (King & Volkov, 2005).

Multidisciplinary Model of Evaluation Capacity Building

This model developed by Preskill and Boyle (2008) is based on various disciplines such as evaluation, organizational learning and change, and adult learning; the objective is to provide a set of guidelines for designing and implementing ECB to maximize its success. Preskill and Boyle (2008) describe their model in the following way: the model is depicted by two circles; the left circle consists of various inner circles that represent the initiation, planning, designing, and implementation of the ECB effort; the outer circle contains the goal of ECB: the development of evaluation knowledge, skills, and attitudes. The content in the right circle recognizes that assumptions, expectations, and motivations about evaluation influence ECB activities and processes and affect the design and implementation of these activities. The inner circles includes 10 ECB strategies that reflect various teaching and learning approaches for helping people develop the knowledge, skills, and attitudes to think evaluatively and to engage in evaluation practice. The left circle is connected to the right circle by a transfer of learning arrow that links the ECB efforts to the requirements to sustain evaluation practice presented in the right circle, including processes, practices, policies, and resources. Both circles are affected by the learning capacity of the organization, leadership, culture, systems and structures, and communication (Preskill & Boyle, 2008).

Evaluation Capacity: A Model

This model was developed by Gibbs, Napp, Jolly, Westover, and Uhl (2002) to “describe beliefs and attitudes related to evaluation and to identify factors influencing evaluation capacity” (p. 261). It is based on interviews conducted with community-based organizations, health departments, and technical assistance providers. According to the authors, the model presents the relationships and interactions of four factors that influence evaluation capacity as well as

three stages of evaluation capacity. The four factors are: (a) funding agency expectations – set a baseline for the amount and type of evaluation activity to be performed; (b) resources – availability of resources such as staff, time, access to external consultants, funding for operational costs, and computer hardware and software to determine evaluation activities; (c) leadership and staff – leaders who believe that evaluation could be used to improve program effectiveness and that strategies could be found to overcome the challenges of evaluation; and (d) evaluation tools and technology – the inclusion of evaluation designs and data collection methods tailored to the specific requirements of the organization (Gibbs, et al., 2002). The three stages of evaluation capacity were described by the authors as: (a) compliance stage – the organization conduct evaluation to the extent required by funding sources and it usually lack the ability or motivation to use the information; (b) investment stage – it goes beyond compliance in the sense that leadership has institutionalized evaluation as a tool for program improvement and commit resources to conduct it; and (c) advancement stage – it goes beyond investment in the sense that organization has broad institutionalized support for evaluation and the use of increasingly sophisticated designs and methods that contribute to a broader understanding of theory and practice (Gibbs, et al., 2002).

An Evaluation Capacity Building Training Framework

This framework developed by Arnold (2006), is based on four strategic methods or components for teaching evaluation: (a) using logic models for program planning providing clear expectations and training and identifying indicators for success and potential points of evaluation at all levels of the model; (b) providing one-on-one consultations for individual evaluation projects rather than general trainings on evaluation; (c) facilitating small-team collaborative evaluations to allow opportunity for exchanges of ideas and methods and diminish feelings of

uncertainty or inadequacy as people learn and practice evaluation skills together; and (d) conducting large-scale multisite evaluation to obtain the benefits of evaluation without having to do all the work. This framework of training and support was developed to increase evaluation capacity, promoting changes to the evaluation culture of the 4-H program, and provide a linear step-by-step framework to develop evaluation capacity with emphasis on individual training Arnold (2006). The author indicates that this framework works well with individuals regardless of the level of evaluation experience of a given person.

The Collaborative Immersion Approach to ECB

This approach developed by Huffman, Thomas, and Lawrenz (2008), “is grounded in social-constructivism learning theory and recognizes that individuals and organizations learn through social, collaborative experiences” (p. 366). The authors explain that this approach focuses on the overall development mechanism to build evaluation capacity for both individuals and organizations through real-world experiences and support the idea of building evaluation capacity by starting with a complex evaluation experience (turning upside down the framework proposed by Arnold in 2006). These authors believe that the growth of individual capacity can take place at the same time as the growth of the organization capacity is occurring; therefore, this experience requires intense immersion of the people in the evaluation process in a way that is not found in typical evaluation methods (Huffman et al., 2008). According to this approach, it is necessary to build evaluation capacity within both individuals and their organization since the beginning of the evaluation, with the expectation that individuals will take what they learned and bring it back to their organizations (Huffman et al., 2008). To help individuals increase their knowledge and skills, “...it is important to understand that immersion programs are intentionally designed to place participants in complex situations that provide independent thought and

encourage growth and development with support” (Huffman et al., 2008, p. 365). The authors emphasize also the importance of having a team composed of individuals within the organization and external evaluators with expertise in evaluation, as a fundamental piece of this approach. The intention of this approach is to build capacity of individuals to participate in, conduct, and use evaluation, while developing the capacity of the organization, and its ultimate goal is to build capacity for future work in evaluation (Huffman et al., 2008).

Integrated Evaluation Capacity Building Model

This model was originally developed by Labin, Duffy, Meyers, Wandersman, and Lesesne (2012), including key elements presented in the ECB theoretical and empirical literature, in order to guide a research synthesis of the empirical literature on ECB. As a result of this research synthesis the model was modified and expanded (Labin, 2014). Labin et al. (2012) explain that they used a logic model format to organize and depict the integrated ECB model including needs, activities, and outcomes and that their intention was to develop a model including the most important ECB activities and processes. As presented by Labin (2014), the three main areas of the updated model are: (a) need/reasons – existing needs and strengths including internal and external factors; (b) activities and mediators – ECB strategies and effectiveness of the strategies mediated by implementation, evaluation, and organizational capacity factors; and (c) outcomes – at the individual, organizational, and program level. This is a linear, two-dimensional model that displays direct and indirect relationships represented by two types of arrows and represents as the author indicates, a “snapshot at a point in time” to envision the interactions in the ECB process (Labin, 2014).

Evaluation Capacity Building/Development in Nonprofit Organizations

The majority of nonprofit organizations depend on internal evaluation (Carman & Fredericks, 2008). Thus, evaluations conducted in nonprofit organizations are a great example of the extent to which in-house evaluators, which usually include the organizational leadership, program managers, and other personnel that serve many roles to get the job done, have also become responsible for major aspects of evaluation in their organizations (Fitzpatrick et al., 2011; Lambur, 2008).

Funders who support nonprofit organizations are realizing that stronger organizations could have better program impact (Connolly & York, 2002). According to these authors, systematic evaluations help organizational leadership articulate the value of their work and compare the effectiveness of different capacity building activities. As a result, nonprofit organizations are experiencing an increased external (e.g., funders, policy makers) and internal demand to conduct evaluations to demonstrate their effectiveness and efficiency (Bozzo, 2002; Carman, 2009; Connolly & York, 2002). This situation generates concerns for these organizations because they are required to provide evidence regarding accountability, how they are reaching the established organizational goals, produce information that assist the organization in providing better services (Connolly & York, 2002; King & Volkov, 2005), and support their competitiveness for funding (Bozzo, 2002).

Newcomer (2004) indicates that nonprofit organizations are behind responding to the increased demand for evaluation because their capacity is not developing at the expected pace. Carman and Fredericks (2010) present similar findings saying that many nonprofit organizations have difficulties gathering evaluation information. They explain that many of these organizations report the need of additional resources and technical assistance to improve their

evaluation capacity. According to Bozzo (2002), nonprofit organizations have to face many challenges when evaluating their programs, including not only the availability of resources (e.g., human and financial) and evaluation skills (e.g., lack of skills to approach an evaluation and produce findings) but also the design of evaluation (e.g., selection of an appropriate evaluation approach, lack of in-house capacity to undertake data collection and analysis), and the nature of nonprofit work (e.g., compliance with different evaluation requirements from multiple funding sources, projects conducted in partnership).

In an effort to improve the evaluation capacity of nonprofit organizations, various authors (Bozzo, 2002; Carman & Fredericks, 2010; Newcomer, 2004) propose evaluation capacity building as part of the solution before undertaking evaluation activities, and acknowledge the importance of establishing sustainability of capacity building efforts. Specifically, Bozzo (2002) explains that some of the priorities in nonprofit organizations to build evaluation capacity should consist of “fostering collaboration, encouraging dialogue among the players, meeting resource and skill needs, addressing methodological challenges, ensuring flexibility in evaluation, and building a feedback loop into evaluation” (p.82).

Internal Evaluation

Internal evaluations are defined by Scriven (1991) as “those done by project staff, even if they are special evaluation staff -- that is, even if they are external to the production/writing/teaching/- service part of the project” (p. 197), and are usually part of formative evaluation efforts. Love (1991) defines internal evaluation as “the process of using staff members who have the responsibility for evaluating programs or problems of direct relevance to an organization’s managers” (p. 2). According to Fitzpatrick et al. (2011), internal evaluations are the ones conducted by organizational employees; and external evaluations are those conducted by

outsiders. More recently, the definition proposed by Volkov describes internal evaluation as “a comprehensive and context-dependent system of intraorganizational processes and resources for implementing and promoting evaluation activities” (Volkov & Baron, 2011, p. 102).

Internal evaluators have firsthand knowledge of the organization history, policies, procedures, personnel, and the decision making-process (Fitzpatrick et al., 2011; Love, 1991, 2005). They also have organizational memory, knowledge of organizational programs, context, and operations, knowledge and understanding of organizational culture, and ability for follow-up or remind others of results more frequently (Beere, 2003; Conley-Tyler, 2005; Fitzpatrick et al., 2011; Sonnichsen, 1999). According to Beere (2003), such knowledge is fundamental to effective ECB. Internal evaluators “can help to create evaluation capacity within their organization” (p. 88) by using strategies found in ECB models/approaches/frameworks that fit the organization’s leadership and culture (Baron, 2011).

Internal evaluators can serve as a source of information to assist organizational leadership in determining the effectiveness and efficiency of the organization and the achievement of organizational goals and objectives (Baron, 2011; Fitzpatrick et al., 2011; Sonnichsen, 1999). Also, they can help organizations to learn (Sonnichsen, 1999), encourage greater use of evaluation information (Love, 1991), communicate relevant evaluation information in a timely fashion (Love, 2005), and improve the quality of programs and services (Baron, 2011). Also, internal evaluators can initiate evaluations when necessary (Baron, 2011). Love (1991) explains that the aforementioned benefits extend “beyond individuals and groups to the organization as a whole” (p. 5), making the development of an internal evaluation system an investment for any organization.

Debates about the efficacy of internal against external evaluators or evaluations, usually turn into controversies in an effort to demonstrate which approach is better (Sonnichsen, 1999). Internal and external evaluations are different and the differentiation between them, to determine which one makes sense, depends on more than simple criteria (Mathison, 2011). This type of issue presents limitations in the sense that the controversy does not lead to explorations of the potential of each approach to determine which one matches better the information needs of the organization (Sonnichsen, 1999). The choice of any of these forms of evaluation should consider the purpose of the evaluation and who is in the best position to conduct the evaluation (Volkov, 2011). Mathison (1991) explains that internal and external evaluations face some similar issues, but there are some unique problems and issues of internal evaluations because internal evaluators operate from within organizations. She states that, “it is clear that organizational factors are paramount in the definition and conduct of internal evaluation” (p. 164).

The internal evaluation process is essential for organizational survival as Love (1983, 1991, 2005) indicates, because crucial evaluative information can be obtained for regulatory compliance, program improvement, strategic planning, accountability or any other need, particularly when resources are limited, building an evaluation culture. Minnett (1999) concurs with Love that internal evaluation provide invaluable information for program development, but she goes further emphasizing that to facilitate learning and change, the evaluation results should be available and fully understood by stakeholders. She explains that, “the key linkage between individual and organizational learning occurs when evaluation findings are shared with members, and they engage in a shared reflection about practice” (p. 354), promoting the integration of evaluation as part of the job. According to Baron (2011), some reasons that make the internal evaluation process usually easier for larger organizations than for smaller organizations are the

availability of resources, funding opportunities from numerous sources, exposure of the organizational personnel to evaluation, and access to evaluation training and expertise.

Fitzpatrick et al. (2011) and Lambur (2008) point out that active support from the organizational leadership and roles for internal evaluator that are clearly delineated are essential conditions for successful internal evaluations. Specifically, internal evaluation systems are intended to influence and contribute to decision making within the organization (Fitzpatrick et al., 2011; Love, 1983). Also, they are important to achieve the legislative evaluation and reporting requirements, to do a job well done (Volkov, 2011). Thus, the personnel in charge of conducting internal evaluations need the active support of leaders within the organization, not only to conduct their roles effectively but also to incorporate evaluation into the decision-making process of the organization (Fitzpatrick et al., 2011; Lambur, 2008).

When organizations are aware of the value and benefits of internal evaluations, they have the potential to initiate evaluations within their organizations simply because it makes good sense and not because it is imposed (Sonnichsen, 1999). If evaluation is perceived as a component of an organizational information processing system, then the evaluation procedures become part of the standard procedures of the organization (Mathison, 1991). Moreover, if the organizational staff identifies themselves as primary stakeholders for evaluation results, they are more open to engage in the process of conducting evaluations (Lambur, 2008).

Love (1991) suggests a model to explain how internal evaluation tends to grow. The model includes six developmental stages of organizational evaluation capacity. The following are the stages described by Love: stage 1: Ad hoc evaluation – managers recognize evaluation as a valuable tool to support decision making but it is hardly ever used and the evaluation capability is primitive and inefficient; stage 2: Systematic internal evaluation – internal evaluation

information is mainly descriptive, the system is well documented and formal processes are in place, but some organizational structures may need to be revised to facilitate the flow of evaluation feedback; stage 3: Goal evaluation – organization begins to design an internal evaluation capability that is relevant to their mission and goals and internal evaluation information is mostly comparative; stage 4: Effectiveness evaluation – organization begins an effort to evaluate the effectiveness of the evaluation and its programs and evaluative information begins to be managed as a corporate resource; stage 5: Efficiency evaluation – criteria are established for measuring efficiency of the organization and programs in converting inputs into outputs and managers take more responsibility for the use of evaluative information across the entire organization; and stage 6: Strategic benefit – senior managers tend to see internal evaluation information as essential strategic tool and they are guided by their organization’s mission and philosophy.

Evaluation Checklists

A checklist is described as a tool or instrument consisting of a detailed list of factors, activities, tasks, items, elements, properties, aspects, components, criteria, or dimensions for convenient checking and reference to perform a certain task (Scriven, 2005, 2007; Stufflebeam, 2001). According to Hales, Terblanche, Fowler, and Sibbald (2008) a checklist is “an organized tool that outlines criteria of consideration for a particular process. It functions as a support resource by delineating and categorizing items as a list--a format that simplifies conceptualization and recall of information” (p. 22). Specifically, an evaluation checklist “serves as a reminder of the processes, procedures, and tasks that need to be addressed in an evaluation” (JCSEE, 2011, p.287). It is intended for guiding a project to success and/or judging its merit and worth, however, it is not a rigid set of rules (Stufflebeam, 2001).

Checklists are commonly seen in many areas including evaluation, as cognitive aids to guide users through accurate task completion (Hales et al., 2008). According to Scriven (2005, 2007) there are different types of checklists, including the laundry list, strongly/weakly sequential, iterative, diagnostic, and criteria of merit checklists (see Table 1 for a description of each one), but one nondefinitional function in common among them is that of being a mnemonic device. The criteria of merit checklists are probably the most important type for evaluation purposes and it is used when ratings or scores are necessary (Scriven, 2005, 2007).

Table 1

Types of Checklists^a

Type	Description
Laundry list	The order of the items does not affect validity, but it is essential to classify the items in the right category.
Strongly/weakly sequential	The order of the items and overall flow is important and affects validity.
Iterative	The order of the items is of somewhat important and requires repeated reviews of each checkpoint to obtain valid results.
Diagnostic	Items are formatted based on flowcharts and often leads to broad, causal conclusions.
Criteria of merit	Commonly used for evaluative purposes and when rating or scores and completeness of criteria is essential to draw conclusions.

^aBased on Scriven, 2005, 2007.

Certain requirements for the criteria or checkpoints in the criteria of merit checklist are suggested by Scriven (2007) including the following: the list should be complete or very close to complete (i.e., avoid significant omissions) and concise, the items should be contiguous (nonoverlapping), and the criteria should be clear. Scriven (2007) also warns the developers of checklists to keep in mind the balance between ease of use and length, taking into consideration brevity which is desirable and clarity which he described as essential. Specifically, “The design

of good checklists should be related to ease of recall and understanding as well as comprehensiveness and ease of implementation” (Scriven, 1991, p. 80). An evaluation checklist should identify relevant dimensions of value (Scriven, 1991, 2007), makes clear the criteria against which the activity or performance will be measured when evaluating something in particular, aids the evaluator remember important criteria, and enhances the objectivity, credibility, and reproducibility of the assessment (Stufflebeam, 2000).

Additional guidelines are provided by Stufflebeam (2000) to develop evaluation checklists. The guidelines are divided into 12 main checkpoints including: (a) focus the checklist task, (b) make a candidate list of checkpoints, (c) classify and sort the checkpoints, (d) define and flesh out the categories, (e) determine the order of categories, (f) obtain initial reviews of the checklist, (g) revise the checklist content, (h) delineate and format the checklist to serve the intended uses, (i) evaluate the checklist, (j) finalize the checklist, (k) apply and disseminate the checklist, and (l) periodically review and revise the checklist.

A checklist provides “an extremely versatile instrument for determining the quality of many kinds of work, programs, activities, and products and may be used to guide observations or a series of measurement efforts” (Scriven, 1991, p. 80). He says that checklists break down a complex judgment into ones that can be made more reliably; when a checklist is used the probability of omitting a critical factor decreases, consequently, reducing a common cause of low reliability. Thus, checklists are considered valuable and useful in evaluation when carefully developed, validated, and applied (Stufflebeam, 2000). Using a checklist in evaluation as the foundation for assessment and to represent overall merit, worth, or importance of something is perhaps the most important function of this tool (Scriven, 2005). Seven general conclusions about the value of checklists described by Scriven (2005, 2007) are summarized as follows:

- Checklists reduce the chances of forgetting to check something important. In other words, the errors of omission are reduced directly and the errors of commission are reduced indirectly.
- Checklists in general are easier for the lay stakeholder to understand and validate than most theories or statistical analyses.
- Checklists reduce the influence of the halo effect (i.e., tendency of a positive judgment of an evaluation based on a particular positive aspect of it) by forcing the evaluator to consider each relevant dimension of possible merit.
- Checklists reduce the influence of the Rorschach effect (i.e., the tendency to see what one wants to see) by forcing a separate judgment on each dimension and a conclusion based on these judgments.
- The use of valid checklist eliminates the problem of double counting.
- Checklists frequently incorporate a great amount of information about the particular evaluand in an efficient way. Specifically, checklists are a form of knowledge about a domain.
- Generally, evaluative checklists can be developed more easily than theories of the evaluand. Accordingly, checklists can contribute substantially to the improvement of validity, reliability, and credibility of an evaluation, and the useful knowledge about a domain.

Evaluation Capacity Building/Development Checklists

Few evaluation capacity building/development checklists exist. *A Checklist for Building Organizational Evaluation Capacity* was developed by Volkov and King (2007). This checklist

provides a set of guidelines for organizational evaluation capacity building, which implies the incorporation of evaluation routinely into the life of an organization. As the authors point out, it can be a resource for a wide range of stakeholders in organizations seeking to increase their long-term capacity to conduct and use program evaluations in everyday activities. The *Evaluation Capacity Development: A Diagnostic Guide and Action Framework* was developed by the World Bank (Mackay, 1999). This document presents detailed checklists of key issues to consider in preparing an action plan for developing evaluation capacities and it was created to assist governments and development agencies to develop a national or sectoral (when national is not feasible) evaluation system. The *Institutionalizing Evaluation Checklist* developed by Stufflebeam (2002) identifies steps that need to be taken in order to install or strengthen, assess, and maintain an evaluation unit supporting the development of evaluation capacity within an organization. It provides a picture of evaluation for those stakeholders who are not familiar with the scope of an evaluation system.

Summary

Evaluation capacity building/development is a recent conceptual development. (King & Volkov, 2005; Trevisan, 2002), and has been recognized as “a multidimensional construct involving maximizing potential in different areas” (Kirsh et al., 2005, p. 235). ECD involves not only the expertise needed to conduct high-quality evaluations but also promotes an organizational culture in which evaluation is part of the everyday work (Beere, 2005). Various definitions of ECB and ECD have emerged based on different research contexts resulting in varying conceptualizations; however, they share some common characteristics or traits (Naccarella et al., 2007; Tseng, 2011). The definition that guided this study is the one proposed

by ECDG (2009a), who defines ECD as “the capacity of putting in place structures which support evaluation efforts within an organization”

ECD as an emergent area of practice has experienced an increased interest from governments and organizations that seek to enhance their effectiveness and accountability (Baizerman et al., 2002a; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005). Nevertheless, as ECD matures more research and attention is needed to better comprehend what it takes to successfully sustain this process and to gain an understanding of its practices (Huffman et al., 2008; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008). Even though developing evaluation capacity is complex, King and Volkov (2005) explain that “capacity can be built, slowly and systematically over time, through procedures that make sense even for small organizations” (p. 12).

CHAPTER III

METHODS

This chapter outlines the methods that were used in the study for developing and validating the Organization Readiness for Evaluation Capacity Development (ORECD) Checklist. The purpose of the study, research questions, research design, validation process, and phases of the study are presented. Each of the study phases conducted is described, including the purpose, participants, procedures of data collection and analysis, as applicable. Procedures for the protection of human subjects are described.

Purpose of the Study

This study was conducted for designing and validating a checklist intended to determine the readiness of an organization to develop evaluation capacity. This instrument provided nonprofit organizations (classified as 501(c)(3) public charities) guidelines containing a formal structure to establish the extent to which they are prepared for the development of evaluation capacity, enhancing objectivity and reproducibility of the assessment. The Organizational Readiness for Evaluation Capacity Development (ORECD) Checklist allows organizations, first, to identify their current situation to support the development of internal evaluation capacity; second, to guide the organization in recognizing which areas may be in need of improvement; and third, to determine the progress made by the organization toward readiness for developing evaluation capacity, by revisiting the ORECD checklist when necessary.

The structure of the ORECD checklist consist of various components (e.g., organizational environment, organizational leadership support, resources) identified in the literature, known to contribute to developing evaluation capacity within organizations. A set of items was developed to represent each of these components. The *Standards for Educational and Psychological Testing* [The Standards] (AERA, APA, & NCME, 1999) and the steps for test construction (Crocker & Algina, 2008) were the main guidelines for designing and validating the ORECD checklist.

Research Questions

The following research questions guided this study:

1. To what extent does the Organizational Readiness for Evaluation Capacity Development Checklist integrate elements presented in the literature that support evaluation capacity development/building based on expert review?
 - a. What is the evidence for face validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
 - b. What is the evidence for content validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
2. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist suitable for the intended users as perceived by individuals with formal evaluation training?
3. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist appropriate for the intended uses as perceived by nonprofit organizations?
4. What are the potential positive and negative consequences of the Organization Readiness for Evaluation Capacity Development Checklist?

5. What are the psychometric properties of the Organizational Readiness for Evaluation Capacity Development Checklist in the field study?
- What are the item-to-total correlations for each component?
 - What is the inter-item relationship for each component?
 - What is the relationship between the components of the checklist?
 - What is the relationship between all the items in the checklist?
 - What is the internal consistency for each component?

Research Design

A mixed method research design was utilized in this study. This type of design focuses on collecting, analyzing, and mixing or combining both qualitative and quantitative data in a single study (Creswell & Plano Clark, 2007). In a mixed method research design, the use of quantitative and qualitative techniques can occur either concurrent or sequential (Teddlie & Tashakkori, 2003). The fundamental idea of this research design is that the combination of qualitative and quantitative data provides a better understanding of what is being studied. In other words, it provides more comprehensive evidence. Mixed methods research provides strengths that compensate for the weaknesses of using quantitative or qualitative data alone. By gathering and analyzing both quantitative and qualitative data simultaneously to address the research questions, more thorough information was obtained to enhance the findings of the overall study.

The study was conducted in four phases. Phase 1 comprised the design and construction of the ORECD checklist as well as the development of the Expert Review Form. Phases 2 through 4 involved the validation process of the ORECD checklist. Specifically, Phase 2 consisted of a review of the ORECD checklist by relevant experts in order to obtain face and

content validity evidences. Phase 3 consisted of a pretesting to determine the appropriateness of the wording of the items and format of the ORECD checklist. Phase 4 consisted of a field study in which nonprofit organizations used the ORECD checklist to evaluate their organizations and provided feedback. This phase generated information about the psychometric properties and utility of the ORECD checklist as well as consequential validity evidence. Phases 1 through 4 were conducted separately. Phases 2 through 4 were conducted after the preceding phase was completed, allowing the inclusion of recommendations to the ORECD checklist before conducting the following phase. A complete description of each of these phases is presented in this chapter.

Validation Process

Validity is defined as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (AERA, APA, & NCME, 1999, p. 9). That is, validity is an overall evaluative judgment of the adequacy and appropriateness of the specific inferences or decisions that result from the test scores or other modes of assessment generated by the instrument (McMillan, 2004; Messick, 1995). It has to be evaluated in regard to the purpose of the instrument and how the instrument will be used.

The aim of the validation process is to generate a body of evidence, including strengths and limitations of the instrument, to inform potential users (Sireci, 2009). The validation process requires accumulating evidences from different sources. There is no such thing as different types of validity; validity is a unitary concept. The main sources of validity evidence are test content, response processes, internal structure, relationship to other variables, and consequences of testing (AERA, APA, & NCME, 1999; Gall, Gall, & Borg, 2007; McMillan, 2004).

The evidences can be obtained in several ways depending on the instrument. This also depends on the inferences to be drawn from the scores (Crocker & Algina, 2008). As Sireci notes (2009), it is not clear when sufficient evidence has been obtained because there is a lack of prescriptive rules for what needs to be done. Kane (2009) suggests that the evidence required depends on the proposed interpretations and uses. This author explains, “If the scores have a direct and simple interpretation, little or no evidence would be needed for validation” (p. 46); however when the scores have complex interpretations or are used for multiple purposes, validation can become a difficult process.

Face and Content Validity

The primary sources of evidence that were obtained throughout the design of the ORECD checklist included face and content validity. Face validity evidence implies a casual examination of the instrument items to determine whether they cover the content that the instrument claims to measure, as viewed by laypersons, users, or experts (Allen & Yen, 1979; Gall, Gall, & Borg, 2007). The evidence based on content is related to how adequately the sample of items in the instrument is representative of the domain of content that the instrument is designed to measure (AERA, APA, & NCME, 1999). It is usually established systematically by content experts based on individual, subjective judgment (Allen & Yen, 1979; Crocker & Algina, 2008; Cronbach & Meehl, 1955; Gall, Gall, & Borg, 2007; Messick, 1995; Yaghmale, 2003) and also can be obtained from logical or empirical analysis to determine how much the content represents the content area (AERA, APA, & NCME, 1999). Hence, content validity measures the comprehensiveness and representativeness of the content of an instrument and should be addressed since the beginning of the instrument development (McGartland Rubio et al., 2003; Yaghmale, 2003).

This study depicted early efforts to evaluate the ORECD checklist prior to encouraging its use. Face validity was established by experts. Content validity was established first by the researcher, including the identification of the domain of content through a review of relevant literature, generating the pool of items accordingly, and developing the ORECD checklist. Second, experts examined the content of the ORECD checklist to establish content validity as well including a quantitative and qualitative review.

Consequential Validity

The evidence based on consequences is related to the implementation and use of the instrument and the decisions made based on the interpretations. Some authors (e.g., Nichols & Williams, 2009; Reckase, 1998) recommend obtaining information regarding planned uses and expected consequences in the development process of an instrument. This evidence should include both positive consequences and evidence that unintended consequences are minimal (Brualdi, 1999). Because there are no actual consequences, considering the ORECD checklist is a new instrument, the purpose of the validation process in this study was to inform whether potential benefits or potential unintended consequences of using the instrument are likely to be realized, as advised by the Standards (AERA, APA, & NCME, 1999). This is important because as Levin-Rozalis et al. (2009) indicate, the way in which evaluation is conducted influences the performance of the evaluand. In other words, evaluation has consequential validity.

Consequential validity evidence of the checklist was collected directly from participants during the field study. Also, follow-up interviews were conducted with a sample of the participant organizations. These interviews provided information to establish if the organizations foresee any benefits or issues as a result of using the information from the ORECD checklist to determine their readiness for evaluation capacity development.

Phase 1: Design of the Organizational Readiness for Evaluation Capacity Development Checklist and Expert Review Form

Purpose

This phase entailed the design and construction of the ORECD checklist as shown in Appendix A. Specifically, during this phase, the initial pool of items was written, the ORECD checklist was constructed, and the Expert Review Form was developed. This evaluation form was designed to guide the expert reviewers in the evaluation of the content of the checklist.

Procedure

The ORECD checklist was designed as a self-assessment instrument to be completed by personnel of an organization, preferably by those responsible for conducting evaluations. During the generation of the initial pool of items and construction of the ORECD checklist, an evaluation capacity development expert was consulted and some informal feedback about the content of the items was obtained. The development of the items and the construction of the ORECD checklist by the researcher was based and guided primarily by: (a) the review of the literature regarding evaluation capacity building and evaluation capacity development; (b) the content of the first six tools of the *Evaluation Capacity Development Group Toolkit* (Russon & Russon, 2007); and (c) the review of relevant literature regarding internal evaluation in organizations with emphasis in nonprofit organizations, to provide a solid base to the items. Table 2 contains detailed information about the evidence from the literature reviewed by the time the ORECD checklist was designed to define the components of the ORECD checklist and develop the items. A four-point rating scale (1 = *not at all*, 2 = *to a small extent*, 3 = *to some extent*, 4 = *to a great extent*) was developed to rate each of the items of the ORECD checklist.

Table 2

Literature Reviewed to Design the Components and Items of the Organizational Readiness for Evaluation Capacity Development Checklist

Component	Literature Reviewed
Organizational Environment	Arnold, 2006; Atkinson et al., 2005; Baizerman et al., 2002a; Baron, 2011; Compton, 2009; GAO, 2003; King, 2007; King & Volkov, 2005; Love, 1983; Mackay, 2002; Milstein et al., 2002; Newcomer, 2004; O'Sullivan & O'Sullivan, 1998; Russon & Russon, 2007; Stevenson et al., 2002; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005
Organizational Leadership Support	Adams & Dickinson, 2010; Compton, 2009; Dabelstein, 2003; Forss et al., 2006; García-Iriarte et al., 2011; Hoole & Patterson, 2008; Kapucu, Augustin, & Krause, 2007; Khan, 1998; King, 2002, 2007; King & Volkov, 2005; Majchrzak, 1982; Newcomer, 2004; Russon & Russon, 2007; Stockdill et al., 2002; Taut, 2007b; Taylor-Powell & Boyd, 2008
Knowledge /Skills Development	Adams & Dickinson, 2010; Arnold, 2006; Atkinson et al., 2005; Bozzo, 2002; Compton et al., 2002; Dabelstein, 2003; Duignan, 2003; Forss et al., 2006; Gibbs et al., 2002; Huffman et al., 2008; King, 2002, 2007; King & Volkov, 2005; Kirsh et al., 2005; Milstein et al., 2002; Monroe et al., 2005; O'Sullivan & O'Sullivan, 1998; Preskill, 2008; Preskill & Boyle, 2008; Stevenson et al., 2002; Sonnichsen, 1999; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005
Resources	Adams & Dickinson, 2010; Arnold, 2006; Dabelstein, 2003; García-Iriarte et al., 2011; Gibbs et al., 2002; King, 2002; King & Volkov, 2005; Kirsh et al., 2005; Lennie, 2005; Milstein et al., 2002; O'Sullivan & O'Sullivan, 1998; Preskill & Boyle, 2008; Russon & Russon, 2007; Stockdill et al., 2002
Program Theory	Adams & Dickinson, 2010; Bickman, 1987; Carman, 2007; Compton et al., 2002; Connolly & York, 2002; Frechtling, 2007; GAO, 2003; Gugiu & Rodríguez-Campos, 2007; Kaplan & Garrett, 2005; King, 2002; Lambur, 2008; Monroe et al., 2005; Suárez-Balcázar & Harper, 2003; Taylor-Powell & Boyd, 2008; Volkov & Baron, 2011
Demand for Evaluation	Forss et al., 2006; GAO, 2003; Hauge, 1998; Mackay, 2002; Preskill & Boyle, 2008; Stockdill et al., 2002; Taylor-Powell & Boyd, 2008
Communication	Forss et al., 2006; King, 2002; King & Volkov, 2005; Taylor-Powell & Boyd, 2008
Evaluation Policies and Procedures	Adams & Dickinson, 2010; Dabelstein, 2003; Duignan, 2003; King, 2005; Preskill & Boyle, 2008; Russon & Russon, 2007; Taut, 2007b; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005
External Support	Bozzo, 2002; Carman, 2007; GAO, 2003; King & Volkov, 2005; Monroe et al., 2005; Mott, 2003; Newcomer, 2004; Preskill, 2008; Preskill & Boyle, 2008; Russon & Russon, 2007; Taylor-Powell & Boyd, 2008; Valéry & Shakir, 2005
Incentives	Adams & Dickinson, 2010; Boyle & Lemaire, 1999; Forss et al., 2006; King, 2002; King & Volkov, 2005
Feedback Mechanism	Bozzo, 2002; Khan, 1998; King & Volkov, 2005; Preskill & Boyle, 2008
Evaluation Use	Bozzo, 2002; Dabelstein, 2003; GAO, 2003; Hauge, 1998; King, 2002, 2007; King & Volkov, 2005; Mackay, 2002; Patton, 2008; Preskill & Boyle, 2008; Taut, 2007b

The aim of the literature review from different sources was to provide a conceptualization of evaluation capacity development. It is important to note that the terms evaluation capacity development and evaluation capacity building are used interchangeably in the literature; therefore, the review of the literature included both areas. In addition, “The Guidelines for Developing Evaluation Checklists: The Checklists Development Checklist” (Stufflebeam, 2000) and “The Logic and Methodology of Checklists” (Scriven, 2007) were used as the main sources for designing and formatting the ORECD checklist.

The Expert Review Form was developed by the researcher and included a set of closed-ended questions to determine the relevance of the items to the content area and whether the items were clear or free of ambiguity. Two four-point rating scales to rate the relevance of the item (1 = *not relevant*; 2 = *somewhat relevant*; 3 = *relevant*; 4 = *very relevant*) and the clarity of the items (1 = *not clear*; 2 = *somewhat clear, need revision*; 3 = *clear but need minor revision*; 4 = *very clear*) were used by the experts (see Appendix B). The review form also contained open-ended questions that allowed experts to describe their judgment about the instrument, including recommendations and comments regarding the instruments as a whole, for each component, and for individual items. These questions were used to find out the extent to which the ORECD checklist included sufficient information to determine the organizational readiness to develop evaluation capacity and how the items fit in each component.

Phase 2: Expert Review

Purpose

This phase included a review of the initial ORECD checklist (see Appendix A) by relevant experts. The content of the ORECD checklist was evaluated to obtain feedback about

different aspects, including relevance and clarity of the items and comprehensiveness of the checklist, in order to obtain face and content validity evidences.

Participants

This phase allowed a formal review of the ORECD checklist by five experts in the content area who were selected from a group of potential experts identified by the researcher in view of their expertise using a purposive expert sampling. During the selection process a total of 13 experts were invited at different points in time to obtain feedback from at least five experts. Of those experts that were invited, seven accepted the invitation and five returned their feedback. The expertise was determined based on publication of books and/or journal articles by the experts in one or more of the following areas: evaluation, evaluation capacity development/building, and development of evaluation checklists. The expertise in these areas contributed knowledge and experiences that were useful for the purpose of this phase of the study.

Experts provided information regarding their professional background. All of them indicated having a doctoral degree, 25 to 35 years of experience in the evaluation field, and 15 to 31 years of experience working in evaluation capacity building/development. Three of them also reported experience developing checklists. By the time this phase was conducted, all the experts were currently active in evaluation capacity building/development.

Procedure

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) emphasizes the need of having experts to assess the content of an instrument as part of the validation process. The experts for this phase of the study were contacted and invited to participate by email. The invitation included information about the purpose of the study and the

nature of their role in the study. After their approval to participate as reviewers, the researcher sent by email the initial version of the ORECD checklist, the Expert Review Form, and the Professional Background Questionnaire. Detailed instructions regarding the review of the ORECD checklist and completion of the review form were included. Each expert was instructed to rate the extent to which the items were relevant to the content area and clear or free of ambiguity. Also, they were invited to provide feedback in the form of descriptive data, about their judgment of the ORECD checklist, including the adequacy of the response scale, ways in which the checklist could be improved, ways in which the items could be improved, and any additional recommendation, as needed. The Professional Background Questionnaire was used to obtain information about the experts' professional experience (see Appendix C). The experts were asked to return their review within two weeks. A reminder was sent at the end of the first week and at the end of the second week to request completion of the review, as necessary. In the cases in which an expert did not conduct the review of the checklist, a new expert was invited. This process was repeated until obtaining feedback from five experts.

Data Analysis

The analysis assessed face and content-related evidences through expert review using both quantitative and qualitative procedures. The quantitative analysis included mean score and standard deviation for each of the proposed items across the areas that were evaluated (i.e. relevance and clarity). In addition, the inter-rater reliability among the five experts was estimated using the Intraclass Correlation Coefficient (ICC). According to Shrout and Fleiss (1979), the ICC is the correlation between one measurement on a target and another measurement obtained on that target. The ICC can be used to estimate agreement not only for pairs of measurement but for larger sets of measurements as well (McGraw & Wong, 1996). The

score represents how much similarity or consensus there is in the ratings given by judges. ICC takes into account the differences in ratings for individual segments, along with the correlation between raters.

There are several versions of the ICC, depending on various assumptions (Shrout & Fleiss, 1979; Streiner & Norman, 2008). For the purpose of this study the ICC (2, k) version was estimated. The number 2 inside the parenthesis indicates the statistical model in which all the items are evaluated by all the raters and the raters are a subset of a large set of raters (raters are considered a random factor). The *k* inside the parenthesis indicates that the study focused on the reliability of the mean score provided by a group of raters (e.g., the average rating for *k* judges, the average score for a *k*-item test) and designates how many scores are averaged together to generate each mean (i.e., 5 raters in this study). Accordingly, the ICC (2, 5), a two-way random (Items x Raters) average measures was used. The relative (consistency) and absolute agreement among experts were estimated for the ORECD checklist regarding relevance and clarity. Relative agreement is estimated when systematic differences between raters are irrelevant and absolute agreement is estimated when systematic differences are relevant. Thus, the reliability based on absolute agreement is always lower than for relative agreement (Streiner & Norman, 2008). The quantitative data were analyzed using SPSS 21 and SPSS 22.

The qualitative analytic procedures involved the examination of the responses to the open-ended questions to obtain face and content validity evidences. Specifically, recommendations to make changes to the ORECD checklist, the items, and/or the response scale were analyzed. The coding of the data by question was conducted independently by the researcher and an external reviewer with knowledge in this type of analysis. Percentage of agreement was estimated $[(\text{agreements}/(\text{agreements} + \text{disagreements})) \times 100\%]$.

A revision of individual items followed based on the results of the quantitative and the qualitative pieces. Low mean scores for relevance to the content area, clarity, or any other inconsistency among experts were considered. Decisions about the content of the ORECD checklist were made after the analysis. The ORECD checklist produced as a result of this phase was used in phase 3 and is presented in Appendix D.

Phase 3: Pretesting

Purpose

During this phase, a review of the ORECD checklist was conducted. The ORECD checklist was reviewed by a sample of doctoral graduate students with formal training in evaluation and professional evaluators. The aim was to determine if the wording and format of the checklist was appropriate and identify ways in which the ORECD checklist could be improved (e.g., clarity and readability of items), in order to include further recommendations into it prior to its application in Phase 4.

Participants

The pretesting phase allowed the review of the ORECD checklist by a convenience-purposeful sample of doctoral graduate students with formal training in evaluation and professional evaluators. Formal training in evaluation meant that participants had formal coursework in evaluation, i.e., at least two graduate evaluation courses. Professional evaluators had formal training in evaluation as well. Convenience sampling was used due to the researcher access to this group of people. The sample was purposeful as well because these people had the evaluation knowledge necessary to provide feedback fitting the purpose of the study. During the selection process a total of 13 potential participants, based on availability and accessibility were

invited. Ten participants accepted the invitation and seven returned their feedback about the ORECD checklist. After conducting the analysis of the data, one additional participant returned their feedback. This information was not included in the analysis because Phase 4 was already initiated.

Procedure

The pretesting was conducted after Phase 2 was over to include on the ORECD checklist the recommendations from the experts. Participants for this phase were contacted and invited to take part in the pretesting by email. The invitation included information about the general purpose of the study, the purpose of the pretesting, and the nature of their role in the study. After their approval to participate in the study, the researcher sent by email, the revised version of the ORECD checklist and the Questions for Pretesting (see Appendix E).

Participants were asked to return their review within two weeks. A reminder was sent at the end of the first week and at the end of the second week to request completion of the review, as needed. This phase included, in addition to the ORECD checklist, a set of open-ended questions to capture the feedback from participants in the form of descriptive data. It also included a cover sheet with the instructions and the tasks participants had to complete while reviewing the ORECD checklist.

Data Analysis

The data from the pretesting was analyzed using qualitative procedures and involved the examination of the feedback provided by participants. The coding of the data by question was conducted independently by the researcher and an external reviewer with knowledge in this type of analysis. Percentage of agreement was estimated $[(\text{agreements}/(\text{agreements} + \text{disagreements}))]$

x 100%)]]. Subsequently, a revision of the items on the ORECD checklist followed based on participants' recommendations. The ORECD checklist produced as a result of this phase was used in phase 4 and is presented in Appendix F.

Phase 4: Field Study

Purpose

This phase had two purposes. First, nonprofit organizations were asked to use the ORECD checklist as a self-assessment tool to evaluate their organizations. Second, feedback about the benefits of using the ORECD checklist and potential improvements to it was also requested. In this way, this phase provided data to examine the psychometric properties of the ORECD checklist, evaluate it from the intended users perspective (i.e., nonprofit organizations), and obtain consequential validity evidence.

Participants

The sampling frame for this phase consisted of nonprofit organizations classified as 501(c)(3) public charities, with emphasis on health care, human services, education, youth development/services, and community services, in Hillsborough and Pasco Counties, Florida that were registered with the IRS. Other nonprofit organizations, including 501(c)(3) organizations that offer other type of service(s) as their principal activity (e.g., professional associations, alumni associations, associations of employees, sport clubs, fraternities/sororities, and religious activities); 501(c)(3) public charities that were not registered with the IRS (i.e., very small organizations with gross receipts under \$5,000 and congregations for which there was no information available); 501(c)(3) private foundations, and other tax exempt organizations consisting of 501(c)(4) social welfare; 501(c)(5) labor unions, farm bureaus and more; 501(c)(6)

business league; and 501(c)(7) social and recreational clubs, were excluded from this study. Nonprofit organizations were a priority because they are facing increased pressures to document their effectiveness from their funding sources such as the government, foundations, and other funders, that usually have evaluation requirements. Hillsborough and Pasco Counties were chosen due to their proximity to the researcher and the extensive amount of nonprofit organizations in these locations.

The National Center for Charitable Statistics (NCCS), which is the national clearinghouse of data on the nonprofit sector in the United States, was used as one of the resources to identify the nonprofit organizations for this study. This database available online, contains the nonprofit organizations registered with the IRS. According to data published in December 2011 by the NCCS, Hillsborough County had 3,418 nonprofit organizations registered as 501(c)(3) public charities and Pasco County had 812. An additional database, TaxExemptWorld.com, was used as well to support the identification of the nonprofit organizations and obtain information about their profile. This database contains public record information made available by the federal government. After applying the inclusion criteria to the information in the databases, the sampling frame for this study consisted of 684 nonprofit organizations in Hillsborough and Pasco Counties.

A proportional stratified sample selection was used. The stratification was based on the type of service(s) provided by the nonprofit organizations (i.e., human services, health care, education, youth development/services, community services, and multiple services). First, nonprofit organizations for each stratum were identified including those in Hillsborough and Pasco Counties together. The major principal activity according to the information on the databases, determined the stratum in which the organization belonged. Second, the proportion of

each stratum was estimated (i.e., human services, 41%; health care, 13%; education, 15%; youth development/services, 11%; community services, 5%; and multiple services, 15%). The nonprofit organizations classified as *multiple services* were those with more than one type of service as their principal activity. Third, stratified random samples were selected based on the proportions. For the follow-up interview, nonprofit organizations were selected using simple random sampling from those that indicated their availability to participate in it.

For this phase, confidence intervals to calculate a single Cronbach's alpha were estimated to determine the sample size needed to test this coefficient with the desired precision. A computer program available at StatTools (http://www.stattools.net/SSiz1Alpha_Pgm.php) derived from Bonett's formula 4 (based on probability of Type I Error α , power, the number of items, the anticipated alpha, and a theoretical alpha) published in 2002 was used. A sample size of at least 30 nonprofit organizations was established as an acceptable minimum in terms of precision when conducting the statistical analysis. The analyses for this phase were conducted with data from the 32 organizations that returned their feedback. To obtain this amount of participants it was necessary to select multiple proportional stratified samples.

Overall, 585 (86%) nonprofit organizations were invited to participate in the study: human services ($n=241$), health care ($n=77$), education ($n=86$), youth development/services ($n=63$), community services ($n=31$), and multiple services ($n=87$). Twenty seven research packages did not reach their destination. Assuming that 558 research packages made their destination and that 32 participants returned the documents completed (136 responses received, 102 participants indicated that they were not interested or available to participate, and 2 indicated their availability to participate but did not return the documents), the response rate was estimated at 6%. In addition, 17 participants indicated their availability to participate in the

follow-up interview and five were interviewed, until saturation of the information was reached. All of the participants interviewed were executive directors of their nonprofit organization. Details of each of the samples selected are presented in Table 3 and described next.

Table 3

Composition of the Samples Selected

Type of Service	Participants in Stratified Samples ^a								
	Overall	1	2	3	4	5	6	7	8
Human Services	241	45	31	31	31	31	31	31	10
Health Care	77	14	10	10	10	10	10	10	3
Education	86	16	11	11	11	11	11	11	4
Youth Development	63	12	8	8	8	8	8	8	3
Community Services	31	6	4	4	4	4	4	4	1
Multiple Services	87	17	11	11	11	11	11	11	4
Total ^b	558/585	107/ 110	71/75	74/ 75	73/ 75	66/ 75	72/ 75	73/ 75	22/ 25
Participants	32	8	3	4	6	2	5	2	2
Response Rate	6.0%	7.5%	4.2%	5.4%	8.2%	3.0%	6.9%	2.7%	9.0%

Note. ^aEight samples were selected. ^bTotal indicates the amount of research packages that seemed to reach their destination and the total amount of research packages sent.

The first sample selected included 110 nonprofit organizations: human services ($n=45$), health care ($n=14$), education ($n=16$), youth development/services ($n=12$), community services ($n=6$), and multiple services ($n=17$). Eleven research packages were returned and it was possible to send again eight of them to alternate addresses. Assuming that 107 research packages reached their destination and considering that eight participants returned the documents completed (28 responses were received and 20 participants indicated that they were not interested or available to participate), the response rate was estimated at 7.5%. Three of the eight organizations that completed the documents indicated their interest to participate in the follow-up interview.

The second sample selected included 75 nonprofit organizations: human services ($n=31$), health care ($n=10$), education ($n=11$), youth development/services ($n=8$), community services ($n=4$), and multiple services ($n=11$). Seven research packages were returned and it was possible to send again six of them to alternate addresses from which three were returned again. Assuming that 71 research packages reached their destination and considering that three participants returned the documents completed (17 responses were received, 13 participants indicated that they were not interested or available to participate, and 1 indicated their availability to participate but did not return the documents), the response rate was estimated at 4.2%. Two of the three organizations that completed the documents indicated their interest to participate in the follow-up interview.

The third sample selected included 75 nonprofit organizations: human services ($n=31$), health care ($n=10$), education ($n=11$), youth development/services ($n=8$), community services ($n=4$), and multiple services ($n=11$). Seven research packages were returned and it was possible to send again six of them to alternate addresses. Assuming that 74 research packages reached their destination and considering that four participants returned the documents completed (15 responses were received and 11 participants indicated that they were not interested or available to participate), the response rate was estimated at 5.4%. The four organizations that completed the documents indicated their availability to participate in the follow-up interview.

The fourth sample selected included 75 nonprofit organizations: human services ($n=31$), health care ($n=10$), education ($n=11$), youth development/services ($n=8$), community services ($n=4$), and multiple services ($n=11$). Three research packages were returned and it was possible to send again all of them to alternate addresses from which two were returned again. Assuming that 73 research packages reached their destination and considering that six participants returned

the documents completed (17 responses were received, 10 participants indicated that they were not interested or available to participate, and 1 indicated their availability to participate but did not return the documents), the response rate was estimated at 8.2%. Three of the six organizations that completed the documents indicated their interest to participate in the follow-up interview.

The fifth sample selected included 75 nonprofit organizations: human services ($n=31$), health care ($n=10$), education ($n=11$), youth development/services ($n=8$), community services ($n=4$), and multiple services ($n=11$). Ten research packages were returned and it was possible to send again seven of them to alternate addresses from which six were returned again. Assuming that 66 research packages reached their destination and considering that two participants returned the documents completed (13 responses were received and 11 participants indicated that they were not interested or available to participate), the response rate was estimated at 3.0%. The two organizations that completed and returned the documents indicated they were available to participate in the follow-up interview.

The sixth sample selected included 75 nonprofit organizations: human services ($n=31$), health care ($n=10$), education ($n=11$), youth development/services ($n=8$), community services ($n=4$), and multiple services ($n=11$). Nine research packages were returned and it was possible to send again seven of them to alternate addresses from which one was returned again. Assuming that 72 research packages reached their destination and considering that five participants returned the documents completed (15 responses were received and 10 participants indicated that they were not interested or available to participate), the response rate was estimated at 6.9%. One of the five organizations that completed the documents indicated its availability to participate in the follow-up interview.

The seventh sample selected included 75 nonprofit organizations: human services ($n=31$), health care ($n=10$), education ($n=11$), youth development/services ($n=8$), community services ($n=4$), and multiple services ($n=11$). Four research packages were returned and it was possible to send again two of them to alternate addresses. Assuming that 73 research packages reached their destination and considering that two participants returned the documents completed (23 responses were received and 21 participants indicated that they were not interested or available to participate), the response rate was estimated at 2.7%. One of the two organizations that completed the documents indicated its availability to participate in the follow-up interview.

The eighth sample selected included 25 nonprofit organizations: human services ($n=10$), health care ($n=3$), education ($n=4$), youth development/services ($n=3$), community services ($n=1$), and multiple services ($n=4$). Six research packages were returned and it was possible to send again three of them to alternate addresses. Assuming that 22 research packages reached their destination and considering that two participants returned the documents completed (8 responses were received and 6 participants indicated that they were not interested or available to participate), the response rate was estimated at 9.0%. One of the two organizations that completed the documents indicated its availability to participate in the follow-up interview.

Part of the respondents who expressed that they were not interested in participating provided explanations as to why they did not want to take part of the study. One of the reasons was related to the size of the organization. Being a small organization represented a limitation for some of them. They said in this regard, "I am a very small organization. I do most of this myself, so I don't know how much help I can be," and "We are an extremely small, all volunteer organization. I'm afraid most of the questions are not applicable to us at this time," apparently indicating substantial lack of evaluation capacity.

Some organizations also expressed having limited resources such as staff, time, and funding making their organization unsuitable for the study. Some of their comments were, “Your survey is not applicable to our organization. We do not employ staff,” “Our agency has limited staff and funding at present,” and “We are too small an organization to participate in this project. We don’t have the time.” Additional reasons included, illness in family, no longer be a 501(c)(3), no longer be an active organization, language-barriers, and organization been moved to another state.

Demographics. The demographics of the organizations and information about their evaluation practices provided a description of their characteristics (see Appendix H) and are presented in Table 4. According to data from 30 participants (two did not complete the demographics form), organizations were founded between 1973 and 2011. Almost half of them (43.3%) reported an annual operating budget equal or less than \$25,000 and 26.7% reported an annual operating budget greater than \$1,000,000. These data represent a proxy of the size of the organization, showing that about half of them are small and about a quarter are very large. For the most part, their funds come from foundations, donations, and the state government.

Participants also provided information about the type of employees they have. More than half of the organizations (66.6%) indicated having full-time employees, ranging from 1 to 1000, and almost half (46.4%) reported having part-time employees, ranging from 1 to 258. In addition, 23.3% informed that they only have volunteers in their organizations. In terms of the services offered, the majority of the organizations (70.0%) provide education services as their principal activity, followed by multiple services (63.3%) and youth development/services (56.7%). On the other hand, few participants (10%) indicated that they provide community services.

Table 4

Demographics of Nonprofit Organization Participants

Characteristic	Frequency	Percent
Annual operating budget		
Up to \$25,000	13	43.3
\$25,000 - \$50,000	1	3.3
\$50,001 - \$100,000	2	6.7
\$100,001 - \$250,000	2	6.7
\$250,001 - \$500,000	4	13.3
\$500,001 - \$750,000	0	----
\$750,001 - \$1,000,000	0	----
More than \$1,000,000	8	26.7
Type of employees ^a		
Full-time	20	66.6
Part-time	14	46.6
Volunteers (only)	7	23.3
Services offered ^b		
Education	21	70.0
Health care	10	33.3
Human services	14	46.6
Youth development/services	17	56.7
Community	3	10.0
Multiple services	19	63.3
Funding sources ^c		
Local government	8	26.7
State government	12	40.0
Federal government	9	30.0
Foundation	14	46.6
Donations	13	43.3
Fundraising	5	16.6
Fee for service	4	13.3
Grants	4	13.3
Primarily responsible for the evaluation		
Internal evaluator	9	30.0
External evaluator	6	20.0
Evaluation unit within the organization	5	16.6
Organization staff with evaluation training	5	16.6
Organization staff without evaluation training	9	30.0
Organization does not conduct evaluation	7	23.3
Administrators/Academics	4	13.3
State/Federal government	1	3.3
Evaluation activities of the organization		
Does not conduct evaluation	3	10.0
Does very little evaluation	3	10.0
Does some evaluation	16	53.3
Makes a great effort to evaluate their program(s)	6	20.0
Goes above and beyond to evaluate their program(s)	2	6.7

Note. N=30. ^aAmount of full-time employees 1-1000, amount of part-time employees 1-258. ^bOrganizations with multiple services were also included by individual areas. ^cTwenty organizations have multiple funding sources

Regarding evaluation activities of the organization, more than half (53.3%) reported that they do some evaluation followed by a 20.0% that indicated they make a great effort to evaluate their programs. The primarily responsible of the evaluation is either the internal evaluator (30.0%) or organization staff without evaluation training (30.0%). This finding coincides with what is reported by other authors (Carman & Fredericks, 2008; Fitzpatrick et al., 2011; Lambur, 2008) which indicates that the majority of nonprofit organizations depend on internal evaluation. In addition, about a quarter of the organizations (23.3%) indicated that they do not conduct evaluations. Five out of seven of the organizations that do not conduct evaluation are those that have only volunteer staff.

Additional information was gathered about the amount of people in the organization in charge of completing the ORECD checklist, with 72% indicating that only one person completed the checklist, 16% indicating that two people completed the checklist, and 6% indicating that three people completed the checklist. All of the participants reported that when they completed the checklist they did it thinking about the entire organization and not thinking about a unit or department in the organization.

Procedure

The field study was conducted after Phase 3 was over to include on the ORECD checklist the recommendations from the pretesting. Organizations (i.e., the person(s) who knows more about the organization's evaluation activities or the person(s) designated by the organization) were asked to rate each item on the ORECD checklist based on their perception of their current situation to develop evaluation capacity. In addition, they were invited to provide feedback, in the form of narrative data, about the utility of the ORECD checklist in determining their readiness for developing evaluation capacity and its potential benefits guiding the development

of evaluation capacity. They were also asked to identify ways in which the ORECD checklist could be improved, in order to incorporate further recommendations into it (see Appendix G).

For the purposes of this study, the demographic questions to explore the characteristics of the nonprofit organizations were asked in a separate form (see Appendix H). An additional question in that form explored the availability of the organization to participate in a follow-up interview. The interview examined information about the process followed by the organization in completing the ORECD checklist and issues or benefits they foresee as a result of using this tool (see Appendix I).

A recruitment/consent letter was sent to the nonprofit organizations inviting them to participate in the study. This letter informed organizations about the general purpose of the study, the purpose of the field study, and the nature of their role in the study including the instructions and the tasks they should complete. The organizations also received via priority mail, the ORECD checklist, the Questions for the Field Study form, and the Nonprofit Organization Demographics form. A postage paid return envelope was included with the materials to help defray costs to participants. Organizations were asked to return the completed ORECD checklist and feedback within two weeks. A reminder was sent at the end of the second week to request completion of the checklist and feedback to those organizations that did not respond in the two-week period. To maintain confidentiality, a numeric code was assigned to each organization and included in the demographic questionnaire. A list including the name of the organization and the corresponding code was kept in a separate place and only the researcher had access to this information.

After receiving the ORECD checklist, the Questions for the Field Study form, and the Nonprofit Organization Demographics form completed by the organizations, five of them were

randomly selected from the 17 who indicated their willingness to participate in the follow-up interview. The organizations were selected at different times until saturation of the information was reached. These organizations were contacted either by phone, email, or priority mail to set up a time and date for the face-to-face interview. The interviews were conducted by the researcher and recorded on audio after approval from participants.

Data Analysis

The data gathered from this phase of the study included both quantitative and qualitative procedures. Descriptive statistics, including mean score and standard deviation for each item and each component of the ORECD checklist are provided. Correlations between components, inter-item correlations by component, and correlations between all the items on the ORECD checklist were estimated using Pearson product-moment correlation coefficient. The correlation coefficients were interpreted based on the following guidelines: $r = \pm .10$ to $\pm .29$ = low; $r = \pm .30$ to $\pm .49$ = moderate; $r = \pm .50$ to ± 1.0 = high (Pallant, 2010). Also, Cronbach's alpha was estimated to determine the internal consistency reliability of each of ORECD checklist components. This internal consistency coefficient is appropriate when there is no right or wrong answer to each item and when items contain a range of possible answers. The following guidelines were used to conduct the interpretations of this value: $\alpha \geq .9$ = excellent; $\alpha \geq .8$ = good; $\alpha \geq .7$ = acceptable $\alpha \geq .6$ = questionable; $\alpha \geq .5$ = poor; $\alpha < .5$ = unacceptable (George & Mallery, 2010). Item-to-total correlations were also examined by component along with Cronbach's alpha if item is deleted to determine the contribution of each item to the component. Items with values less than .30 were revised. A low value (less than .3) may indicate that an item is measuring something different (Pallant, 2010). The quantitative data were analyzed using SPSS 22. Listwise deletion

was used to handle missing values to reduce the possibility of problems when different subsets of cases are used for the statistical analysis.

The qualitative analytic procedures involved the transcription and examination of the responses from both the Questions for the Field Study form and the Field Test Interview Protocol form. The coding of the data by question was conducted independently by the researcher and an external reviewer with knowledge and experience in this type of analysis. Percentage of agreement was estimated $[(\text{agreements}/(\text{agreements} + \text{disagreements})) \times 100\%]$. Subsequently, a final revision of the items on the ORECD checklist followed based on participants' feedback and the psychometric properties estimates. The final version of the ORECD checklist produced as a result of this phase is presented in Appendix J.

Protection of Human Subjects

This research study collected data provided by participants including evaluation professionals, doctoral graduate students, and nonprofit organization personnel. There was minimal risk associated with the participation in this study. The study focused on information regarding the characteristics of the ORECD checklist (e.g., item content and clarity, response scale, and utility of the checklist) and included also descriptive information about the demographics of the nonprofit organizations and the professional background of the experts. All the data collection for this study were conducted after the approval by the Institutional Review Board of the University of South Florida (see Appendix K).

Participation in the study was voluntary. Participants could withdraw at any time during the research study without penalty. Participants did not receive monetary compensation for their participation. All participants were informed about the purpose of the research study and the procedures to be followed before their participation, either by email (participants in phase 2 and

3) or mail (participants in phase 4). A waiver of the signature of the consent form was requested to IRB for phases 2, 3, and 4 of the study, considering the way the study was conducted. Returning the completed forms was recognized as consent to participate. For confidentiality purposes, no names were used in phases 2 and 3, and a numeric code was assigned to each organization in phase 4 as a means to identify the organizations that were available to participate in the follow-up interview. All the data gathered are kept in a locked file cabinet in the house of the researcher. Electronic files were saved in an external hard drive property of the researcher. Only the researcher had access to the files. The data will be stored for 5 years after the completion of the study. After that, the paper documents will be shredded and the electronic files will be deleted.

Summary

In this chapter, the phases for the design and validation of a new instrument, the Organizational Readiness for Evaluation Capacity Development Checklist, are presented. The validation process generated a body of information through the collection of qualitative and quantitative data which entailed the accumulation of evidences from different sources. Specifically, the sources to establish evidences included face, content, and consequential validity. The study was designed in such a way that each of the proposed phases depicted early efforts to evaluate and improve the ORECD checklist prior to encouraging its use. Also, information about the psychometric properties and utility of the ORECD checklist was generated.

CHAPTER IV

RESULTS

This chapter presents the results of this study conducted for designing and validating the Organizational Readiness for Evaluation Capacity Development (ORECD) Checklist. The information is organized in the order in which the phases of the study were conducted. This chapter begins by describing how the ORECD checklist was developed followed by the results of the experts' review of the checklist, the pretesting of the checklist, and the field study. The results are then summarized by research question. The following research questions were addressed:

1. To what extent does the Organizational Readiness for Evaluation Capacity Development Checklist integrate elements presented in the literature that support evaluation capacity development/building based on expert review?
 - a. What is the evidence for face validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
 - b. What is the evidence for content validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
2. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist suitable for the intended users as perceived by individuals with formal evaluation training?
3. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist appropriate for the intended uses as perceived by nonprofit organizations?

4. What are the potential positive and negative consequences of the Organization Readiness for Evaluation Capacity Development Checklist?
5. What are the psychometric properties of the Organizational Readiness for Evaluation Capacity Development Checklist in the field study?
 - a. What are the item-to-total correlations for each component?
 - b. What is the inter-item relationship for each component?
 - c. What is the relationship between the components of the checklist?
 - d. What is the relationship between all the items in the checklist?
 - e. What is the internal consistency for each component?

Phase 1: Design of the Organizational Readiness for Evaluation Capacity Development Checklist

The ORECD checklist was designed as a self-assessment instrument to be completed by personnel of an organization, preferably by those responsible for conducting evaluations, as applicable. The development of the items and the construction of the ORECD checklist by the researcher was based on and guided primarily by: (a) the review of the literature regarding evaluation capacity building and evaluation capacity development; (b) the content of the first six tools of the Evaluation Capacity Development Group Toolkit (Russon & Russon, 2007); and (c) the review of relevant literature regarding internal evaluation in organizations, with emphasis in nonprofit organizations, to provide a solid conceptual base to the items of the checklist. During the generation of the initial pool of items, informal feedback about the content of the items and the ECD aspects covered was obtained from an evaluation capacity development expert.

The aim of the literature review from different sources was to provide a conceptualization of evaluation capacity development/building. It is important to note that the terms evaluation

capacity development and evaluation capacity building are used interchangeably in the literature, therefore, the review of the literature included both areas. In addition, The Guidelines for Developing Evaluation Checklists: The Checklists Development Checklist (Stufflebeam, 2000) and The Logic and Methodology of Checklists (Scriven, 2007) were used as the main sources for the designing and formatting of the ORECD checklist.

The initial ORECD checklist designed contained 61 items in 12 components (see Appendix A). Components were created because ECD has been recognized as a multidimensional construct; and its success depends on organizational commitment in different areas (Adams & Dickinson, 2010). Thus, key components known to contribute to developing evaluation capacity within organizations according to the literature, guided the design of the components of the ORECD checklist. These components and the amount of items in each components were: (a) Organizational Environment, 13 items; (b) Organizational Leadership Support, 7 items; (c) Knowledge/Skills Development, 6 items; (d) Resources, 7 items; (e) Program Theory, 3 items; (f) Demand for Evaluation, 4 items; (g) Communication, 3 items; (h) Policies and Procedures, 4 items; (i) External Support, 3 items; (j) Incentives, 4 items; (k) Feedback Mechanism, 3 items; and (l) Evaluation Use, 4 items. A four-point rating scale to rate each of the items (1 = *not at all*, 2 = *to a small extent*, 3 = *to some extent*, 4 = *to a great extent*) was used. Detailed information about how the ORECD checklist was modified after each phase is provided in this chapter.

Phase 2: Expert Review

In this phase feedback from experts about the ORECD checklist was gathered. The feedback included quantitative and qualitative data. The quantitative data analysis consisted of an examination of the closed-ended questions in the Expert Review Form (see Appendix B),

including ratings for relevance and clarity of the items. Inter-rater reliability was estimated using the Intraclass Correlation Coefficient (ICC). The ICC included relative and absolute agreement for relevance and clarity for all the items on the ORECD checklist and was estimated using SPSS 21. The ICC for relevance was .35 for relative agreement and .30 for absolute agreement. The ICC for clarity was .34 for relative agreement and .18 for absolute agreement. The results indicate that the agreement among experts, for the most part, was fair.

Descriptive statistics for each item on the ORECD checklist regarding relevance and clarity are presented in Table 5. The ratings of the items on the ORECD checklist are represented by the mean and the spread of the ratings is represented by the standard deviation of the items. Regarding relevance, the mean of each item (2.40 to 4.00) showed that most of them were rated as *relevant* or *very relevant*. The standard deviation fluctuated from 0.0 to 1.41. This variability in ratings may explain the ICC values obtained. An examination of the individual ratings assigned to each item revealed that only one expert rated 14 items as not relevant. Ten of these items were removed from the ORECD checklist and the other four items were retained because the other experts rated them mostly as *relevant* or *very relevant*. As a result, all the items with a mean rate less than 3.00 for relevance were removed.

In terms of clarity, the mean of each item (1.60 to 3.75) revealed that almost all of them were rated as *somewhat clear*, *need revision* and *clear but need minor revision*. The standard deviation fluctuated from 0.50 to 1.64. This variability in ratings may explain the ICC values obtained and support the modifications made to the retained items. An exploration of the individual ratings assigned to each item revealed that over half of the time experts rated the items as *very clear* or *somewhat clear*, *need revision*.

Table 5

Descriptive Statistics for Relevance and Clarity for Each Original Item

Item	Relevance		Clarity	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1. Organizational environment allows the establishment/strengthening of an internal evaluation system.	3.60	0.55	2.40	0.89
2. Organization is aware of the benefits of internal evaluation.	3.60	0.55	2.60	1.34
3. Role of evaluation is understood by the whole organization (not only the leadership).	3.25	0.50	2.50	1.29
4. Evaluation is valued by the whole organization.	3.25	0.50	2.50	1.29
5. Evaluation is considered a means to provide important information.	3.50	0.58	3.25	0.96
6. Personnel are committed to developing the evaluation capacity of the organization.	3.20	1.30	3.00	0.82
7. Personnel understand how evaluation can contribute to organizational learning.	3.50	1.00	2.75	0.96
8. Personnel are committed to learning from evaluation.	4.00	0.00	3.00	0.82
9. Evaluators are perceived as facilitators.	2.60	1.14	2.50	1.29
10. Evaluators are perceived as a useful resource.	2.60	1.14	2.75	0.96
11. Personnel understand the importance of incorporating evaluation into everyday work practices.	3.20	1.30	2.75	0.96
12. Personnel know how their work relates to evaluation.	3.40	0.89	2.50	1.00
13. External stakeholders (e.g., funders) support the development of evaluation capacity.	3.00	0.82	3.25	0.96
14. Leadership supports the development of internal evaluation capacity.	4.00	0.00	3.20	1.10
15. Leadership has a long-term commitment to support ongoing evaluation capacity development.	3.80	0.45	3.20	1.10
16. Leadership encourages involvement of personnel in the development of evaluation capacity.	3.00	1.00	2.20	1.64
17. Leadership supports practices that integrate evaluation into organizational life.	3.80	0.45	3.00	1.00
18. Leadership supports the integration of evaluation at the decision-making level.	3.67	0.58	2.40	1.14
19. Leadership is committed to the development of an infrastructure to support the development of evaluation capacity.	3.75	0.50	2.50	1.00
20. Leadership is committed to the development of an evaluative learning culture.	3.75	0.50	2.60	1.34
21. Organization is aware of the current evaluation skills/knowledge of the personnel.	2.60	1.14	2.50	1.29
22. There is a mechanism to identify evaluation capacity needs of the personnel.	3.60	0.55	3.00	1.00

Table 5 (Continued) *Descriptive Statistics for Relevance and Clarity for Each Original Item*

Item	Relevance		Clarity	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
23. Organization is committed in providing professional development to increase evaluation skills/knowledge of the personnel.	3.60	0.55	2.00	1.23
24. Training opportunities at different organizational levels are available to support the continuity of evaluation capacity efforts.	2.80	1.30	3.00	1.16
25. Evaluation expertise is available inside the organization.	3.40	1.34	3.50	1.00
26. Personnel with experience conducting evaluations are available.	3.20	1.30	2.75	0.96
27. Organization has sufficient <u>human</u> resources to develop evaluation capacity.	3.80	0.45	3.00	1.00
28. Organization has sufficient <u>financial</u> resources to develop evaluation capacity.	3.80	0.45	3.00	1.00
29. Evaluation budget is a priority for the organization.	4.00	0.00	3.20	0.84
30. Evaluation budget reflects the evaluation needs of the organization.	3.40	1.34	3.75	0.50
31. Evaluation budget is consistent with organization's long term objectives.	3.00	1.41	2.25	1.26
32. Appropriate tools/technologies are available (e.g., computer hardware and software, equipment, materials).	4.00	0.00	2.80	1.10
33. Personnel have sufficient time during the workday to collaborate on evaluation activities.	3.80	0.45	3.00	1.00
34. Personnel are aware of how organizational program(s) work.	3.00	1.41	2.25	0.50
35. Personnel know how to construct logic models.	2.75	1.26	3.00	1.16
36. Logic model(s) is used as an evaluation planning tool (define and clarify what and when to evaluate).	3.00	1.23	3.25	0.96
37. There is demand for information (e.g., on what works, accountability requirements, strategies to increase evaluation use) from external stakeholders.	3.75	0.50	3.00	1.23
38. There is demand to develop evaluation capacity.	3.33	0.58	2.00	1.23
39. There is commitment to develop evaluation capacity.	3.50	.058	2.60	1.34
40. There is a plan to develop evaluation capacity.	3.33	0.58	2.50	1.29
41. An effective organizational communication system is in place.	2.60	1.14	2.50	0.58
42. Organization has communication structures to facilitate the flow of information across the organization (i.e., informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	3.60	0.55	3.20	1.10
43. Organization has communication procedures to manage information (e.g., collection and dissemination).	3.25	0.50	2.40	1.14

Table 5 (Continued) *Descriptive Statistics for Relevance and Clarity for Each Original Item*

Item	Relevance		Clarity	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
44. Organization has evaluation/evaluation capacity development <u>policies</u> in place (e.g., rules to guide evaluation decisions and actions, evaluation training for personnel, ethical considerations, disclosure of information).	3.40	0.89	2.40	1.52
45. Organization has evaluation/evaluation capacity development <u>procedures</u> in place (e.g., handbooks and manuals, sources of and guidelines to obtain technical assistance in evaluation, consultation processes).	3.40	0.89	2.60	1.34
46. Personnel are aware of the evaluation/evaluation capacity development <u>policies</u> .	3.40	0.89	2.60	1.34
47. Personnel are aware of the evaluation/evaluation capacity development <u>procedures</u> .	3.40	0.89	2.60	1.34
48. Organization collaborates with external evaluation experts to support the development of evaluation capacity.	3.20	0.84	2.00	1.00
49. Organization shares evaluation knowledge/skills with other organizations to support the development of evaluation capacity.	2.40	0.89	1.67	0.58
50. Organization employs mechanisms for advancing evaluation capacity (e.g., building networks and relationships, identifying outstanding practices from successful organizations to use them as standards for comparison).	3.40	0.55	2.40	0.89
51. Incentives are available to encourage participation in the development of evaluation capacity (e.g., time and flexibility to incorporate the evaluation process into daily work practices).	3.60	0.55	2.80	1.30
52. Incentives are available to encourage involvement of personnel in evaluation.	3.00	1.16	2.80	1.30
53. Personnel are acknowledged for their contributions to evaluation.	3.60	0.55	3.00	1.00
54. Personnel understand how evaluation can contribute to their individual learning (e.g., development of valuable, lifelong skills).	3.50	0.58	2.75	1.50
55. Organization has a feedback mechanism.	3.00	0.00	1.60	0.89
56. Personnel are receptive to feedback from others.	3.25	0.50	2.20	0.84
57. Feedback mechanism is part of the decision-making process.	3.25	0.50	2.20	0.84
58. Organizational leadership supports the use of evaluative information.	4.00	0.00	2.80	1.64
59. Dissemination of evaluation results is promoted.	3.40	0.55	1.80	0.84
60. Personnel are able to use evaluation results (e.g., for planning, decision-making).	3.80	0.45	2.80	1.30
61. Evaluation results are used for continuous improvement of the organization.	3.80	0.45	3.00	1.41

Note. Relevance: 1 = not relevant; 2 = somewhat relevant; 3 = relevant; 4 = very relevant. Clarity: 1 = not clear; 2 = somewhat clear, need revision; 3 = clear but need minor revision; 4 = very clear.

Qualitative data analysis of the open-ended questions in the Expert Review Form was conducted as well. The data were coded using codes developed by the researcher based on the experts' responses to the questions. Two sets of codes were generated and used for the analysis. One set was utilized to analyze the questions in Part 1 of the form and the other set of codes was used to analyze the questions in Part 2. The questions in Part 1 asked experts to comment about the extent to which the items in a component fit the content of that component. The data were coded according to the following codes: *items fit*; *items fit but revise*; and *move item to another component*. The questions in Part 2 asked experts to provide feedback about the ORECD checklist, including appropriateness of the checklist as a measure of the elements necessary for ECD, potential changes to the checklist, clarity of the items, appropriateness of the response scale, and items that should have been revised, added or deleted. The data were coded according to the following codes: *item changes*, *checklist modifications*, *response scale appropriateness*, and *checklist appropriateness*. The coding of responses for each question was independently conducted by two researchers. The agreement between raters was 97% for Part 1 and 93% for Part 2.

Experts' responses provided in Part 1, regarding the content fit of the items by component, were coded as *items fit* when the group of items was suitable for the corresponding component. These are, "They relate to resource [component] issue" and "Make sense to me." Responses coded as *items fit but revise* were those with recommendations for improvement to the group of items in a particular component. There was more feedback pertaining to this code than for any other code. Examples are, "They fit, but several need more clarity about what the construct looks like – there's probably more jargon than necessary. I would try to write these in a way that respondents can understand in their everyday work practices," "I think these items

could be reduced,” “These items could be clearer and more specific,” “These items seem more appropriate as part of the ECB work, not as part of ‘readiness’ to engage in ECB,” “These items are essential but, do they belong as worded as part of ‘readiness’?,” and “This section needs more work –greater clarity in terms of type of communication system/processes/procedures you are talking about.”

Responses coded as *move item to another component* include, for instance, “If items are only about external stakeholders, recommended that you move them to ‘I’ [External Support],” “[items] 39 and 40 seem to go in ‘B’ [Organizational Leadership Support] or ‘H’ [Policies and Procedures],” and “Is this [Feedback Mechanism] part of the communication system, ‘G’?” Overall, these results indicated the need to do modifications to the content of several items and changes to some of the components.

It is important to note that not all the experts offered feedback about each component. A summary by component is given here to provide a better understanding of their feedback. In regard to the Organizational Environment component they indicated that the items fit but they could be reduced because some of them overlap. For the Organizational Leadership Support, Knowledge/Skills Development, Resources, and Program Theory components, experts stated that the items seem to be related to the component but that some of them needed clarification such as adding an example or more information and rewording or making the items more specific. In the feedback provided for the Demand for Evaluation component, suggestions were mainly about moving some items to another component or revise them considering both the internal and external demand for evaluation, not only the external demand. For the Communication component, recommendations were about improving the clarity of the items. They also recommended moving the items in the Feedback Mechanism component to the Communication

component. The recommendation was adopted and the items were placed under Communication considering that feedback is an important part of any communication system. In the Policies and Procedures and Incentives components, experts indicated that items seemed to fit and suggested to add examples to some of the items to make them clearer. Experts also stated that the items in the Evaluation Use component needed rewording to align them better to readiness. For the items in the External Support component, they indicated that some of them seemed to be part of another item in the same component.

Experts' responses provided in Part 2 were coded as *item changes* when items needed some type of modification in terms of grammar, rewording, revisions due to jargon, adding or deleting words, and clarification of the content. Example responses include, "Reduce the amount of jargon (evaluation lingo) and make more specific to evaluation," "[items] need clarification," "I found the feedback mechanism items unclear; not sure exactly what you were meaning," "Get rid of all the passive voice; add the actors who need to do certain things," "Change personnel to staff," and "I think items would be better and more clear if they were more specific. You might accomplish this by including 'as demonstrated by'."

Responses coded as *checklist modifications* were those with recommendations to add items, merge items, modify components, revise the order of the components, and align the content to readiness. Examples consist of, "Collapse [items] 7 and 8," "Move [item 13] to 'I component,'" "Are the twelve categories in a meaningful order?" "I wonder if it would be better to limit the checklist to items relevant to the readiness stage of ECB?," "Component 'G' - include: Organizational members are committed to sharing information; There is a transparent and understood process for sharing information," "...you don't use the idea of a written plan for ECB. It's implicit in a couple of places, but I find if people don't have an actual, physical plan

that ECB can easily go by the wayside. That's one possible addition," and "I haven't marked any [items for deletion], but if some of the ones I've marked unclear can't be reliably fixed, then I might want to eliminate those."

The *response scale appropriateness* code was used to codify responses about the response scale of the ORECD checklist. Overall, the responses were positive including, "It seems okay to me," "Scale is ok – but it really depends on what you're trying to find out," and "Yes [response scale is appropriate]." Responses coded as *checklist appropriateness* were those with feedback about the usefulness of the ORECD checklist as a measure of the ECD elements. Examples are, "It could be useful if the items are revised and made more accessible and clear," "It seems that you have included all the main components" and "In general, yes [checklist appear to be a good measure of the elements necessary for ECD]."

Even though the questions in Part 1 and Part 2 were different, there was some overlapping of the feedback offered. As a result of the examination of the recommendations provided by experts, several changes were made to the ORECD checklist. Some modifications were made to the items in order to better denote readiness to develop evaluation capacity. Also, *personnel* was changed to *staff*, as recommended, after an analysis of the meaning of these words. It was determined that the meaning of *staff* better represents the message of the ORECD checklist. This word is defined as, "the officers chiefly responsible for the internal operations of an institution or business; the personnel who assist a director in carrying out an assigned task" (Merriam-Webster's Collegiate online dictionary, n.d.). On the other hand, *personnel* is defined as, "a body of persons usually employed, as in a factory, office, or organization" (Merriam-Webster's Collegiate online dictionary, n.d.).

The feedback from experts guided several decisions made about the ORECD checklist. These decisions were about deleting, retaining, modifying or adding items; moving items to another component; and maintaining or collapsing the components of the ORECD checklist. Also, some items were added as example to other items. Table 6 contains the specific action taken with each item after the experts' feedback.

Table 6

Action Taken with the ORECD Checklist Items After the Expert Review

Original Item	Action	New Item
1. Organizational environment allows the establishment/ strengthening of an internal evaluation system.	M	Internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).
2. Organization is aware of the benefits of internal evaluation.	M	Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).
3. Role of evaluation is understood by the whole organization (not only the leadership).	E	
4. Evaluation is valued by the whole organization.	E	
5. Evaluation is considered a means to provide important information.	M	There is a general understanding of how evaluation can provide important information to the organization.
6. Personnel are committed to developing the evaluation capacity of the organization.	M	Organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.
7. Personnel understand how evaluation can contribute to organizational learning.	M	There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.
8. Personnel are committed to learning from evaluation.	M	There is a general commitment to learning from evaluation throughout the organization
9. Evaluators are perceived as facilitators.	D	
10. Evaluators are perceived as a useful resource.	D	

Table 6 (Continued) *Action Taken with the ORECD Checklist Items After the Expert Review*

Original Item	Action	New Item
11. Personnel understand the importance of incorporating evaluation into everyday work practices.	D	
12. Personnel know how their work relates to evaluation.	M	Staff is aware of how their work relates to evaluation.
13. External stakeholders (e.g., funders) support the development of evaluation capacity.	R	
14. Leadership supports the development of internal evaluation capacity.	R	
15. Leadership has a long-term commitment to support ongoing evaluation capacity development.	M	Leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).
16. Leadership encourages involvement of personnel in the development of evaluation capacity.	M	Leadership is committed to encouraging evaluation capacity development activities.
17. Leadership supports practices that integrate evaluation into organizational life.	M	Leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.
18. Leadership supports the integration of evaluation at the decision-making level.	E	
19. Leadership is committed to the development of an infrastructure to support the development of evaluation capacity.	E	
20. Leadership is committed to the development of an evaluative learning culture.	M	Leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).
21. Organization is aware of the current evaluation skills/knowledge of the personnel.	E	
22. There is a mechanism to identify evaluation capacity needs of the personnel.	M	There is a plan to identify the strengths and weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).
23. Organization is committed in providing professional development to increase evaluation skills/knowledge of the personnel.	M	Leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.
24. Training opportunities at different organizational levels are available to support the continuity of evaluation capacity efforts.	D	

Table 6 (Continued) *Action Taken with the ORECD Checklist Items After the Expert Review*

Original Item	Action	New Item
25. Evaluation expertise is available inside the organization.	M	There is staff with evaluation expertise/experience conducting evaluations inside the organization to support the ECD process.
26. Personnel with experience conducting evaluations are available.	E	
27. Organization has sufficient <u>human</u> resources to develop evaluation capacity.	M	There are strategies in place to be able to access sufficient <u>human</u> resources for the ECD process.
28. Organization has sufficient <u>financial</u> resources to develop evaluation capacity.	M	Organization has committed <u>financial</u> resources to develop evaluation capacity.
29. Evaluation budget is a priority for the organization.	R	
30. Evaluation budget reflects the evaluation needs of the organization.	D	
31. Evaluation budget is consistent with organization's long term objectives.	D	
32. Appropriate tools/technologies are available (e.g., computer hardware and software, equipment, materials).	M	There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.
33. Personnel have sufficient time during the workday to collaborate on evaluation activities.	M	There is a plan to provide staff sufficient time during the workday to work on evaluation activities.
34. Personnel are aware of how organizational program(s) work.	M	Staff has a common understanding of <u>how</u> organizational program(s) work (for example, there is a logic model or any other graphical representation of the program(s) and the expected outcomes).
35. Personnel know how to construct logic models.	D	
36. Logic model(s) is used as an evaluation planning tool (define and clarify what and when to evaluate).	E	
37. There is demand for information (e.g., on what works, accountability requirements, strategies to increase evaluation use) from external stakeholders.	M	There is demand for evaluative information from <u>external</u> stakeholders (for example, on what works, accountability requirements, and strategies to increase evaluation use).
38. There is demand to develop evaluation capacity.	M	There is demand for evaluative information from <u>internal</u> sources.
39. There is commitment to develop evaluation capacity.	D	
40. There is a plan to develop evaluation capacity.	M	There is a written plan about how to develop evaluation capacity.

Table 6 (Continued) *Action Taken with the ORECD Checklist Items After the Expert Review*

Original Item	Action	New Item
41. An effective organizational communication system is in place.	M	There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).
42. Organization has communication structures to facilitate the flow of information across the organization (i.e., informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	M	There are communication <u>structures</u> to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).
43. Organization has communication procedures to manage information (e.g., collection and dissemination).	M	There are communication <u>procedures</u> to manage/share information (such as, collection, dissemination, and disclosure of information).
44. Organization has evaluation/evaluation capacity development <u>policies</u> in place (e.g., rules to guide evaluation decisions and actions, evaluation training for personnel, ethical considerations, disclosure of information).	M	There is a plan in place to develop ECD <u>policies</u> (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).
45. Organization has evaluation/evaluation capacity development <u>procedures</u> in place (e.g., handbooks and manuals, sources of and guidelines to obtain technical assistance in evaluation, consultation processes).	M	There is a plan in place to develop ECD <u>procedures</u> (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).
46. Personnel are aware of the evaluation/evaluation capacity development <u>policies</u> .	M	There is a plan to make ECD <u>policies</u> accessible to all staff.
47. Personnel are aware of the evaluation/evaluation capacity development <u>procedures</u> .	M	There is a plan to make ECD <u>procedures</u> accessible to all staff.
48. Organization collaborates with external evaluation experts to support the development of evaluation capacity.	M	Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.
49. Organization shares evaluation knowledge/skills with other organizations to support the development of evaluation capacity.	D	
50. Organization employs mechanisms for advancing evaluation capacity (e.g., building networks and relationships, identifying outstanding practices from successful organizations to use them as standards for comparison).	M	There is a plan to establish mechanisms for advancing the development of evaluation capacity (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).

Table 6 (Continued) *Action Taken with the ORECD Checklist Items After the Expert Review*

Original Item	Action	New Item
51. Incentives are available to encourage participation in the development of evaluation capacity (e.g., time and flexibility to incorporate the evaluation process into daily work practices).	M	Incentives are available to encourage staff participation in the development of evaluation capacity (for example, time and flexibility people need to incorporate evaluation into the everyday work of the organization).
52. Incentives are available to encourage involvement of personnel in evaluation.	D	
53. Personnel are acknowledged for their contributions to evaluation.	M	There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.
54. Personnel understand how evaluation can contribute to their individual learning (e.g., development of valuable, lifelong skills).	M	Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).
55. Organization has a feedback mechanism.	M	There is a feedback mechanism (action or means used to modify the ECD process as a result of information received) to facilitate learning within the evaluation capacity development process.
56. Personnel are receptive to feedback from others.	M	Leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.
57. Feedback mechanism is part of the decision-making process.	D	
58. Organizational leadership supports the use of evaluative information.	M	Leadership supports the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).
59. Dissemination of evaluation results is promoted.	M	Leadership promotes the dissemination of evaluation results (to inform staff, as appropriate, about them).
60. Personnel are able to use evaluation results (e.g., for planning, decision-making).	M	Leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).
61. Evaluation results are used for continuous improvement of the organization.	E	

Note. D = deleted; E = added as an example to another item; M = modified to add content/examples to the item and/or reword it; R = retained without changes.

Of the three original items in the Feedback Mechanism component, one of them was deleted (Item 57). Regarding individual items, Item 3 and Item 4 were deleted as main items and added as examples to Item 2. Similarly, Item 19 was added as an example to Item 15 and Item 18 and Item 61 were added as examples to Item 60. Item 21 was added as an example to Item 22, information from Item 26 was added to Item 25, and Item 36 was added as an example to Item 34. Also, Item 13 was moved to the External Support component. Altogether, three items were retained as they were originally designed, 39 items were modified, 11 items were deleted and eight of them were added as examples to other items, and five new items were added. The purpose of incorporating new items, as per experts recommendation, was to include additional elements that seem to be part of the evaluation capacity development construct that were not previously considered (see Table 7). No changes were made to the four-point rating scale on the ORECD checklist (i.e., 1 = *not at all*, 2 = *to a small extent*, 3 = *to some extent*, 4 = *to a great extent*) because the feedback received from experts was positive.

Table 7

New Items Added to the ORECD Checklist

Component	Item
Organizational Environment	There is organizational stability (for example, organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and staff turnover is low).
Knowledge/Skills Development	There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work).
Resources	There is a plan for securing additional fiscal resources to develop evaluation capacity, as needed.
Program Theory	Organizational program(s) goals/objectives are well defined. Staff has a common understanding about what organizational program(s) do.

The introduction and directions of the ORECD checklist were also revised. The introduction was modified to add a definition of evaluation capacity development to facilitate stakeholders understanding of the concept. The introduction includes this additional sentence: The goal of evaluation capacity development (ECD) is to put in place and sustain the components that support program evaluation efforts within the organization. The adjustments to the directions were made to clarify stakeholders what to keep in mind when completing the ORECD checklist and indicate how to interpret the results. The modified set of directions are: For each of the following statements, circle the response that best describe the current situation of your organization, indicating the extent to which it is present within the organization. After completing the checklist, review the statements marked as “to a small extent” and “not at all”. These are the areas of improvement that need to be prioritized to support the development of evaluation capacity.

As a result of the changes made to the original ORECD checklist after the experts review, the revised version used in Phase 3 contained 47 items across 11 components. These components and the amount of items in each one were: (a) Organizational Environment, 8 items; (b) Organizational Leadership Support, 5 items; (c) Knowledge/Skills Development, 4 items; (d) Resources, 6 items; (e) Program Theory, 3 items; (f) Demand for Evaluation, 3 items; (g) Communication, 5 items; (h) Policies and Procedures, 4 items; (i) External Support, 3 items; (j) Incentives, 3 items; and (k) Evaluation Use, 3 items.

Phase 3: Pretesting

In this phase of the study, doctoral graduate students with formal training in evaluation and professional evaluators provided feedback about the ORECD checklist. A qualitative data analysis of the responses in the Questions for Pretesting form was conducted. The data were

coded using codes developed by the researcher based on participants' responses to the questions. The codes were used to analyze the feedback provided and make modifications to the ORECD checklist accordingly. The coding of the responses for each question was independently conducted by two researchers. The agreement between raters was 95%.

The questions asked participants to provide feedback on the appropriateness of the wording and format of the ORECD checklist and identify ways in which the checklist could be improved (e.g., clarity and readability of items). The data were coded according to the following codes: *review item*, *grammatical changes*, *review response scale*, *clear checklist*, *appropriate flow*, and *appropriate format*.

Responses were coded as *review item*, when recommendations were given about the wording of the item, the examples included as part of the item, or the component in which the item was located. Example responses include, "All B items should you add the words 'The organizational' to the beginning of each item," "I suggest that the statement A.1 should be divided into sub items and each item ask about one specific element," "A.1: The example you provided to support the statement consists of multiple elements. The staff might have one or two but not the other(s)," and "How if C.14 becomes B.14? I still see that statement 14 asking about Leadership Support (mainly supporting improving skills or knowledge)." Other responses were coded as *grammatical changes* when they consisted of minor grammar suggestions such as adding "the, an, or in order to," removing the underline of certain words, and removing or exchanging a word. The following is an example, "A.6 should there be the word 'the' at the beginning of the sentence." Responses coded as *review response scale* include for example, "I suggest asking the user to respond to a scale such as strongly agree, agree..." and "Consider changing the Likert scale. Maybe: strongly agree, agree...not sure."

Responses coded as *clear checklist* were those with positive feedback about the ORECD checklist or specific items. Examples are, “I can say that I can think of nothing to improve upon what you’ve done. The checklist is clear, concise, and thorough,” “Yes [items are clearly worded]; nothing [needs to be changed to improve the items],” “I think you created a great checklist that is going to be very useful for organizations to assess ECD,” and “All in all, I think you have developed a great tool.” Responses coded as *appropriate flow* were those indicating that the items were ordered appropriately. For instance, “The items flow clearly and logically,” “Yes, items flow nicely,” and “Under each category, the items are listed in a logical order.” Responses were coded as *appropriate format* when participants indicated that the physical format is appropriate for its purpose. For example, “Yes, the physical format is okay” and “The tool is well organized and easy to read.”

As a result of the recommendations provided by participants some changes were made to the ORECD checklist. The directions of the ORECD checklist were expanded including a sentence about the purpose of the examples presented as part of some of the items. Also, Item 14, which was originally in the Knowledge/Skills Development component was moved to the Organizational Leadership Support component. No items were removed from the ORECD checklist as a result of the feedback. Only modifications were recommended and most of them were adopted. The specific changes made to the items can be found in Table 8.

Recommendations for changes to the response scale were not adopted because the purpose of the ORECD checklist is to guide organizations to describe the extent of their current situation regarding readiness for evaluation capacity development and guide them in the establishment of areas for improvement. Using a Likert scale to determine the level of

agreement with a statement was considered nonfunctional because it would not provide information of the extent to which an organization has in place for each specific item.

Table 8

Action Taken with the ORECD Checklist Items After the Pretesting

Original Item	Action	New Item
1. Internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).	M	The internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).
2. Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).	R	
3. There is a general understanding of how evaluation can provide important information to the organization.	R	
4. There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.	R	
5. There is a general commitment to learning from evaluation throughout the organization.	M	There is a general commitment to learning from evaluation (process and/or results) throughout the organization.
6. Organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.	R	
7. Staff is aware of how their work relates to evaluation.	R	
8. There is organizational stability (for example, organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and staff turnover is low).	M	There is organizational stability (for example, the organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and the staff turnover is low).
9. Leadership supports the development of internal evaluation capacity.	M	The organizational leadership supports the development of internal evaluation capacity.

Table 8 (Continued) *Action Taken with the ORECD Checklist Items After the Pretesting*

Original Item	Action	New Item
10. Leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).	M	The organizational leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).
11. Leadership is committed to encouraging evaluation capacity development activities.	M	The organizational leadership is committed to encouraging evaluation capacity development activities.
12. Leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.	M	The organizational leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.
13. Leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).	M	The organizational leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).
14. Leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.	M	The organizational leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.
15. There is a plan to identify the strengths and weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).	R	
16. There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work).	M	There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work vs. traditional lectures).
17. There is staff with evaluation expertise/experience conducting evaluations inside the organization to support the ECD process.	M	There is staff with evaluation expertise/experience conducting evaluations inside the organization in order to support the ECD process.
18. There are strategies in place to be able to access sufficient <u>human</u> resources for the ECD process.	M	There are strategies in place to be able to access sufficient human resources for the ECD process.
19. Organization has committed <u>financial</u> resources to develop evaluation capacity.	M	Organization has committed financial resources to develop evaluation capacity.

Table 8 (Continued) *Action Taken with the ORECD Checklist Items After the Pretesting*

Original Item	Action	New Item
20. There is a plan for securing additional fiscal resources to develop evaluation capacity, as needed.	M	There is a plan for securing additional fiscal resources to develop evaluation capacity.
21. Evaluation budget is a priority for the organization.	M	An evaluation budget is a priority for the organization.
22. There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.	R	
23. There is a plan to provide staff sufficient time during the workday to work on evaluation activities.	R	
24. Organizational program(s) goals/objectives are well defined.	M	The goals/objectives of the organizational program(s) are well defined.
25. Staff has a common understanding about <u>what</u> organizational program(s) do.	M	Staff has a common understanding about what organizational program(s) do.
26. Staff has a common understanding of <u>how</u> organizational program(s) work (for example, there is a logic model or any other graphical representation of the program(s) and the expected outcomes).	M	Staff has a common understanding of how organizational program(s) work (for example, there is a logic model or other graphical representation of the program(s) and the expected outcomes).
27. There is demand for evaluative information from <u>external</u> stakeholders (for example, on what works, accountability requirements, and strategies to increase evaluation use).	M	There is demand for evaluative information from external stakeholders, such as funders (for example, on what works, accountability requirements, and strategies to increase evaluation use).
28. There is demand for evaluative information from <u>internal</u> sources.	M	There is demand for evaluative information from internal sources.
29. There is a written plan about how to develop evaluation capacity.	R	
30. There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).	R	
31. There are communication <u>structures</u> to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	M	There are communication structures to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).
32. There are communication <u>procedures</u> to manage/share information (such as, collection, dissemination, and disclosure of information).	M	There are communication procedures to manage/share information (such as, collection, dissemination, and disclosure of information).

Table 8 (Continued) *Action Taken with the ORECD Checklist Items After the Pretesting*

Original Item	Action	New Item
33. There is a feedback mechanism (action or means used to modify the ECD process as a result of information received) to facilitate learning within the evaluation capacity development process.	R	
34. Leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.	M	The organizational leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.
35. There is a plan in place to develop ECD <u>policies</u> (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).	M	There is a plan in place to develop ECD policies (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).
36. There is a plan in place to develop ECD <u>procedures</u> (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).	M	There is a plan in place to develop ECD procedures (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).
37. There is a plan to make ECD <u>policies</u> accessible to all staff.	M	There is a plan to make ECD policies accessible to all staff.
38. There is a plan to make ECD <u>procedures</u> accessible to all staff.	M	There is a plan to make ECD procedures accessible to all staff.
39. External stakeholders (for example, funders) support the development of evaluation capacity.	R	
40. Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.	R	
41. There is a plan to establish mechanisms for advancing the development of evaluation capacity (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).	R	
42. Incentives are available to encourage staff participation in the development of evaluation capacity (for example, time and flexibility people need to incorporate evaluation into the everyday work of the organization).	M	Incentives are available to encourage staff participation in the development of evaluation capacity (for example, allotted time and flexibility for people to incorporate evaluation into the everyday work of the organization).

Table 8 (Continued) *Action Taken with the ORECD Checklist Items After the Pretesting*

Original Item	Action	New Item
43. There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.	R	
44. Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).	R	
45. Leadership promotes the dissemination of evaluation results (to inform staff, as appropriate, about them).	M	The organizational leadership promotes the dissemination of evaluation results (to inform staff about them as appropriate).
46. Leadership supports the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).	M	The organizational leadership promotes the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).
47. Leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).	M	The organizational leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).

Note. M = modified to add content/examples to the item and/or reword it; R = retained without changes.

As a result of the changes made to the ORECD checklist following the recommendations of the pretesting phase, the revised version used in Phase 4 contained 47 items across 11 components. These components and the amount of items in each one were: (a) Organizational Environment, 8 items; (b) Organizational Leadership Support, 6 items; (c) Knowledge/Skills Development, 3 items; (d) Resources, 6 items; (e) Program Theory, 3 items; (f) Demand for Evaluation, 3 items; (g) Communication, 5 items; (h) Policies and Procedures, 4 items; (i) External Support, 3 items; (j) Incentives, 3 items; and (k) Evaluation Use, 3 items.

Phase 4: Field Study

This phase of the study included quantitative and qualitative data gathered from the nonprofit organizations participants in the field study. The quantitative data analysis consisted of an examination of the psychometric properties of the ORECD checklist using SPSS 22. The qualitative data included an analysis of the open-ended questions in the Questions for Field Study form as well as the open-ended questions in the semi-structured interview in the Field Test Interview Protocol. The data were coded using codes developed by the researcher that emerged from participants' responses to the questions. Two sets of codes were generated and used for the analysis. One set was utilized to analyze the questions in the Questions for Field Study form and the other set of codes was used to analyze the questions in the Field Test Interview Protocol. The coding of responses for each question was independently conducted by two researchers. The agreement between raters was 92% for the responses in the Questions for Field Study form and 94%, for the responses from the Field Test Interview Protocol.

Psychometric Properties of the ORECD Checklist

Descriptive statistics, including mean scores and standard deviations are provided for each component and item on the ORECD checklist. The responses to the items within each component were averaged to create composites which were used to estimate the mean and standard deviation of each component. Cronbach's alpha was used as the reliability coefficient to determine the internal consistency of the components of the ORECD checklist. The data analysis also examined item-to-total correlation and alpha if item is deleted of each item by component. Moreover, correlations between components, inter-item correlations for each component, and inter-item correlations for all the items on the ORECD checklist were estimated

as well. As in any checklist, in which the items (checkpoints) represent relevant dimensions of the construct, none of the items required reverse coding.

Descriptive statistics and internal consistency. The data were screened through an examination of skewness and kurtosis for each component which yielded adequate results. Also, the mean, standard deviation, and Cronbach's alpha for each component were estimated (see Table 9). Any value equal to or greater than .70 was considered adequate (George & Mallery, 2010). Cronbach's alpha for the Organizational Environment, Organizational Leadership Support, Resources, Program Theory, Communication, Policies and Procedures, Evaluation Use, Knowledge/Skills Development and Incentives components were greater than .70. The alpha for the Demand for Evaluation and External Support components, were below .70.

Table 9

Psychometric Properties of the ORECD Checklist Components

Component	Number of Items	<i>M</i>	<i>SD</i>	<i>α</i>
Organizational Environment	8	3.24	0.65	.91
Organizational Leadership Support	6	3.40	0.69	.93
Knowledge/Skills Development	3	2.61	0.94	.85
Resources	6	2.00	0.83	.90
Program Theory	3	3.50	0.70	.90
Demand for Evaluation	3	2.33	0.80	.61
Communication	5	3.06	0.73	.90
Policies and Procedures	4	2.13	1.01	.93
External Support	3	2.66	0.71	.66
Incentives	3	2.33	1.01	.86
Evaluation Use	3	2.90	1.07	.94

Note. *N*=32. Means represent the evaluation capacity for each component.

The mean, standard deviation, item-to-total correlation, and alpha if item if deleted for all the items on the OREDC checklist are presented in Table 10. The following components showed

high item-to-total correlation: Program Theory (.78 to .88), Policies and Procedures (.77 to .90), and Evaluation Use (.84 to .91). The item-to-total correlation for the Organizational Environment (.48 to .85), Organizational Leadership Support (.66 to .88), Knowledge/Skills Development (.54 to .87), Resources (.59 to .85), Communication (.57 to .90), and Incentives (.69 to .77) components were all moderate to high. Lastly, the item-to-total correlation of the Demand for Evaluation (.22 to .60) and the External Support (.36 to .54) components were low to moderate. Specifically, in the Demand for Evaluation component, the weakest item of the group, with an item-to-total correlation of .22, was Item 29. In the External Support component, Item 41, with an item-to-total correlation of .36 was the weakest of the group. Therefore, an exploration of the content of these two items was conducted.

Table 10

Psychometric Properties of the ORECD Checklist Items

Component	<i>M</i>	<i>SD</i>	Item-to-Total Correlation	Cronbach's Alpha if Item is Deleted
Organizational Environment				
1. The internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).	3.41	0.71	.82	.89
2. Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).	3.28	0.77	.85	.89
3. There is a general understanding of how evaluation can provide important information to the organization.	3.42	0.72	.78	.90
4. There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.	3.37	0.83	.73	.90
5. There is a general commitment to learning from evaluation (process and/or results) throughout the organization	3.31	0.86	.77	.90

Table 10 (Continue) *Psychometric Properties of the ORECD Checklist Items*

Component	<i>M</i>	<i>SD</i>	Item-to- Total Correlation	Cronbach's Alpha if Item is Deleted
6. The organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.	2.75	1.14	.60	.92
7. Staff is aware of how their work relates to evaluation.	3.10	0.80	.81	.89
8. There is organizational stability (for example, the organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and the staff turnover is low).	3.31	0.74	.48	.92
Organizational Leadership Support				
9. The organizational leadership supports the development of internal evaluation capacity.	3.59	0.67	.74	.93
10. The organizational leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).	3.37	0.83	.84	.92
11. The organizational leadership is committed to encouraging evaluation capacity development activities.	3.34	0.79	.88	.91
12. The organizational leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.	3.44	0.76	.84	.92
13. The organizational leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).	3.34	0.87	.88	.91
14. The organizational leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.	3.28	0.85	.66	.94
Knowledge/Skills Development				
15. There is a plan to identify the strengths/weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).	2.50	1.08	.87	.63
16. There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work vs. traditional lectures).	2.62	1.07	.76	.75

Table 10 (Continue) *Psychometric Properties of the ORECD Checklist Items*

Component	<i>M</i>	<i>SD</i>	Item-to- Total Correlation	Cronbach's Alpha if Item is Deleted
17. There is staff with evaluation expertise/experience conducting evaluations inside the organization in order to support the ECD process.	2.72	1.09	.54	.95
Resources				
18. There are strategies in place to be able to access sufficient human resources for the ECD process.	2.16	1.02	.85	.87
19. Organization has committed financial resources to develop evaluation capacity.	1.84	0.92	.71	.89
20. There is a plan for securing additional fiscal resources to develop evaluation capacity.	1.81	0.98	.80	.88
21. An evaluation budget is a priority for the organization.	1.78	0.87	.59	.91
22. There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.	2.19	1.18	.82	.88
23. There is a plan to provide staff sufficient time during the workday to work on evaluation activities.	2.25	1.14	.69	.90
Program Theory				
24. The goals/objectives of the organizational program(s) are well defined.	3.62	0.61	.83	.87
25. Staff has a common understanding about what organizational program(s) do.	3.56	0.72	.88	.80
26. Staff has a common understanding of how organizational program(s) work (for example, there is a logic model or other graphical representation of the program(s) and the expected outcomes).	3.31	0.93	.78	.92
Demand for Evaluation				
27. There is demand for evaluative information from external stakeholders, such as funders (for example, on what works, accountability requirements, and strategies to increase evaluation use).	2.68	1.11	.49	.42
28. There is demand for evaluative information from internal sources.	2.53	1.08	.60	.23
29. There is a written plan about how to develop evaluation capacity.	1.81	1.00	.22	.77
Communication				
30. There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).	3.16	0.81	.90	.84

Table 10 (Continue) *Psychometric Properties of the ORECD Checklist Items*

Component	<i>M</i>	<i>SD</i>	Item-to-Total Correlation	Cronbach's Alpha if Item is Deleted
31. There are communication structures to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	3.06	0.88	.83	.86
32. There are communication procedures to manage/share information (such as, collection, dissemination, and disclosure of information).	3.03	0.78	.73	.88
33. There is a feedback mechanism (action or means used to modify the ECD process as a result of information received) to facilitate learning within the evaluation capacity development process.	2.53	1.05	.75	.88
34. The organizational leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.	3.50	0.80	.57	.91
Policies and Procedures				
35. There is a plan in place to develop ECD policies (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).	2.19	1.06	.77	.93
36. There is a plan in place to develop ECD procedures (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).	1.97	1.00	.79	.93
37. There is a plan to make ECD policies accessible to all staff.	2.19	1.20	.90	.89
38. There is a plan to make ECD procedures accessible to all staff.	2.16	1.17	.90	.89
External Support				
39. External stakeholders (for example, funders) support the development of evaluation capacity.	2.68	0.95	.51	.50
40. Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.	3.06	0.88	.54	.46
41. There is a plan to establish mechanisms for advancing the development of evaluation capacity (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).	2.25	0.95	.36	.70

Table 10 (Continue) *Psychometric Properties of the ORECD Checklist Items*

Component	<i>M</i>	<i>SD</i>	Item-to-Total Correlation	Cronbach's Alpha if Item is Deleted
Incentives				
42. Incentives are available to encourage staff participation in the development of evaluation capacity (for example, allotted time and flexibility for people to incorporate evaluation into the everyday work of the organization).	2.22	1.13	.75	.79
43. There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.	2.31	1.20	.77	.77
44. Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).	2.47	1.08	.69	.84
Evaluation Use				
45. The organizational leadership promotes the dissemination of evaluation results (to inform staff about them as appropriate).	2.81	1.15	.84	.95
46. The organizational leadership promotes the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).	2.94	1.13	.89	.91
47. The organizational leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).	2.94	1.11	.91	.89

Note. $N=32$.

Correlations. Correlations between components of the ORECD checklist are displayed in Table 11. These correlations ranged from .31 to .83. All the correlations except two of them were significant (most of them at the $p < .01$ level). In addition, all the correlations were positive and more than half of them were high. The correlation between Knowledge/Skills Development and External Support ($r=.31$) and between Program Theory and Policies and Procedures ($r=.34$), even though were moderate in magnitude, are non-significant. The highest correlation found was between Organizational Environment and Communication ($r=.83$) followed by Organizational Environment and Organizational Leadership Support ($r=.78$) and three other

correlations ($r=.71$), including Organizational Leadership Support and Communication, Knowledge/Skills Development and Communication, and Incentives and Evaluation Use.

Table 11

Correlations of the ORECD Checklist Components

Component	OE	OL	KS	RE	PT	DE	CM	PP	ES	IN	EU
OE	---										
OL	.78**	---									
KS	.62**	.56**	---								
RE	.67**	.58**	.56**	---							
PT	.62**	.42*	.48**	.44*	---						
DE	.58**	.54**	.43*	.55**	.43*	---					
CM	.83**	.71**	.71**	.64**	.68**	.47**	---				
PP	.59**	.57**	.37*	.57**	.34	.60**	.47**	---			
ES	.55**	.47**	.31	.49**	.50**	.49**	.41*	.35*	---		
IN	.61**	.64**	.41*	.62**	.49**	.38*	.64**	.55**	.44*	---	
EU	.57**	.52**	.37*	.62**	.67**	.41*	.65**	.40*	.58**	.71**	---

Note. $N=32$. OE = Organizational Environment; OL = Organizational Leadership Support; KS = Knowledge/Skills Development; RE = Resources; PT = Program Theory; DE = Demand for Evaluation; CM = Communication; PP = Policies and Procedures; ES = External Support; IN = Incentives; EU = Evaluation Use.
* $p < .05$. ** $p < .01$.

A summary of the inter-item correlations by component according to their magnitude is displayed in Table 12 and Table 13 presents all the correlations. The inter-item correlations by component were all positive and ranged from .05 to .99. All the inter-item correlations for the following components were high and significant at the $p < .01$ level: Organizational Leadership Support ($r=.54$ to .85), Knowledge/Skills Development ($r=.55$ to .92), Program Theory ($r=.79$ to .85), Policies and Procedures ($r=.72$ to .99), Incentives ($r=.63$ to .74), and Evaluation Use ($r=.81$ to .89). In the Policies and Procedures component it was found that the correlation between Items 37 and 38 was almost perfect ($r=.99$). Also, the inter-item correlations in the Resources ($r=.45$ to .79) and Communication ($r=.45$ to .88) components were significant, most of them at the $p<.01$ level. These correlations were mainly high and only a few were moderate.

Table 12

Amount of Correlations by Component According to Their Magnitude

Component	# of Items	Low +.10 to .29	Moderate +.30 to .49	High > +.50
Organizational Environment	8	1 ^a	3 ^a , 6 ^b	18 ^b
Organizational Leadership Support	6			15 ^b
Knowledge/Skills Development	3			3 ^b
Resources	6		4 ^b	11 ^b
Program Theory	3			3 ^b
Demand for Evaluation	3	2 ^a		1 ^b
Communication	5		1 ^b	9 ^b
Policies & Procedures	4			6 ^b
External Support	3		2a, 1 ^b	
Incentives	3			3 ^b
Evaluation Use	3			3 ^b

Note. ^aAmount of non-significant correlations. ^bAmount of significant correlations.

In the Organizational Environment component the inter-item correlations ($r=.22$ to $.81$) were for the most part significant. More than half of the items in this component were highly correlated. The remainder items in this component had moderate associations except for the correlation between Item 4 and Item 8 which resulted in a low association. It was also found that even though Item 8 has moderate correlations with Item 3, Item 5, and Item 7, these were non-significant. Similarly, in the Demand for Evaluation component, of the three inter-item correlations ($r=.05$ to $.64$) only one was significant and high in magnitude. This correlation was between Item 27 and Item 28. The correlations between Item 27 and Item 29 and between Item 28 and Item 29 were non-significant and low in magnitude. Also, in the External Support component the inter-item correlations ($r=.32$ to $.47$) were for the most part non-significant. The non-significant correlations were between Item 39 and Item 41 and between Item 40 and Item 41, even though they were moderate. Only the correlation between Item 39 and Item 40 was significant and moderate.

Table 13

Intercorrelations of the ORECD Checklist Items

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
OE																								
1	---																							
2	.79 [^]	---																						
3	.63 [^]	.76 [^]	---																					
4	.55 [^]	.66 [^]	.81 [^]	---																				
5	.73 [^]	.78 [^]	.70 [^]	.77 [^]	---																			
6	.68 [^]	.60 [^]	.43 [*]	.46 [*]	.43 [*]	---																		
7	.72 [^]	.73 [^]	.74 [^]	.59 [^]	.67 [^]	.63 [^]	---																	
8	.41 [*]	.43 [*]	.33	.22	.36	.47 [*]	.37	---																
OL																								
9	.64 [^]	.59 [^]	.55 [^]	.69 [^]	.62 [^]	.52 [^]	.51 [^]	.26	---															
10	.60 [^]	.72 [^]	.33	.50 [^]	.61 [^]	.66 [^]	.51 [^]	.29	.62 [^]	---														
11	.74 [^]	.73 [^]	.51 [^]	.67 [^]	.72 [^]	.61 [^]	.55 [^]	.48 [^]	.70 [^]	.82 [^]	---													
12	.66 [^]	.77 [^]	.50 [^]	.63 [^]	.71 [^]	.57 [^]	.63 [^]	.39 [*]	.59 [^]	.82 [^]	.84 [^]	---												
13	.63 [^]	.70 [^]	.42 [*]	.58 [^]	.54 [^]	.62 [^]	.39 [*]	.39 [*]	.76 [^]	.83 [^]	.85 [^]	.79 [^]	---											
14	.57 [^]	.63 [^]	.49 [*]	.42 [*]	.43 [*]	.73 [^]	.57 [^]	.59 [^]	.54 [^]	.58 [^]	.65 [^]	.62 [^]	.62 [^]	---										
KS																								
15	.47 [*]	.61 [^]	.32	.41 [*]	.44 [*]	.55 [^]	.36	.35	.41 [*]	.59 [^]	.49 [*]	.48 [*]	.60 [^]	.42 [*]	---									
16	.48 [^]	.55 [^]	.23	.38 [*]	.43 [*]	.49 [^]	.30	.41 [*]	.34	.54 [^]	.48 [*]	.49 [^]	.59 [^]	.38 [*]	.92 [^]	---								
17	.44 [*]	.34	.34	.34	.29	.57 [^]	.47 [*]	.37	.49 [*]	.33	.34	.18	.37	.36	.61 [^]	.55 [^]	---							
RE																								
18	.62 [^]	.48 [*]	.32	.24	.41 [*]	.56 [^]	.59 [^]	.46 [*]	.28	.49 [*]	.57 [^]	.49 [*]	.40 [*]	.59 [^]	.38 [*]	.45 [*]	.45 [*]	---						
19	.37	.30	.22	.14	.20	.42 [*]	.43 [*]	.30	.32	.34	.35	.37	.25	.57 [^]	.26	.25	.38 [*]	.73 [^]	---					
20	.58 [^]	.47 [*]	.36	.32	.40 [*]	.55 [^]	.48 [^]	.34	.40 [*]	.40 [*]	.55 [^]	.39 [*]	.44 [*]	.57 [^]	.48 [*]	.47 [*]	.52 [^]	.71 [^]	.61 [^]	---				
21	.42 [*]	.21	.31	.33	.32	.50 [^]	.55 [^]	.15	.36	.23	.33	.35	.23	.36	.14	.17	.50 [*]	.48 [^]	.45 [*]	.62 [^]	---			
22	.66 [^]	.59 [^]	.35	.35	.39 [*]	.77 [^]	.49 [^]	.43 [*]	.39 [*]	.64 [^]	.65 [^]	.54 [^]	.59 [^]	.62 [^]	.59 [^]	.59 [^]	.61 [^]	.79 [^]	.58 [^]	.75 [^]	.50 [^]	---		
23	.48 [*]	.47 [*]	.41 [*]	.32	.36	.60 [^]	.55 [^]	.59 [^]	.26	.43 [*]	.51 [^]	.55 [^]	.38 [*]	.71 [^]	.17	.28	.21	.65 [^]	.49 [^]	.52 [^]	.48 [^]	.62 [^]	---	

Table 13 (Continued) *Intercorrelations of the ORECD Checklist Item*

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
PT																								
24	.65 [^]	.46*	.47*	.41*	.31	.63 [^]	.53 [^]	.62 [^]	.49*	.23	.47*	.35	.41*	.61 [^]	.41*	.43*	.54 [^]	.38*	.29	.40*	.43*	.50 [^]	.45*	---
25	.68 [^]	.49*	.52 [^]	.39*	.54 [^]	.48*	.57 [^]	.66 [^]	.48*	.22	.48*	.34	.31	.58 [^]	.33	.31	.45*	.33	.21	.36	.38*	.33	.43*	.83 [^]
26	.51 [^]	.48*	.55 [^]	.38*	.36	.56 [^]	.57 [^]	.66 [^]	.45*	.26	.45*	.35	.38*	.66 [^]	.41*	.31	.59 [^]	.29	.25	.35	.41*	.41*	.42*	.79 [^]
DE																								
27	.25	.18	.12	.24	.15	.19	.29	.13	.51 [^]	.23	.28	.39*	.28	.23	-.04	.02	.13	.10	.24	-.06	.24	-.01	.20	.37
28	.29	.46*	.15	.22	.23	.41*	.44*	.38*	.38*	.54 [^]	.43*	.59 [^]	.47 [^]	.33	.34	.40*	.34	.40*	.45*	.21	.34	.38*	.42*	.31
29	.37	.40*	.38*	.38*	.38*	.48 [^]	.38*	.25	.36	.33	.39*	.36	.30	.50*	.41*	.34	.35	.37	.60 [^]	.66 [^]	.54 [^]	.42*	.34	.35
CM																								
30	.69 [^]	.74 [^]	.58 [^]	.51 [^]	.62 [^]	.52 [^]	.66 [^]	.48*	.44*	.45*	.56 [^]	.61 [^]	.55 [^]	.53 [^]	.59 [^]	.60 [^]	.35	.39*	.11	.39*	.29	.47*	.40*	.57 [^]
31	.57 [^]	.60 [^]	.48 [^]	.36	.41*	.48*	.50 [^]	.65 [^]	.37	.36	.55 [^]	.50 [^]	.54 [^]	.52 [^]	.55 [^]	.61 [^]	.37	.46*	.17	.41*	.28	.52 [^]	.46*	.68 [^]
32	.55 [^]	.51 [^]	.38*	.21	.32	.59 [^]	.56 [^]	.42*	.22	.21	.29	.36	.31	.66 [^]	.48*	.47*	.43*	.52 [^]	.38*	.53 [^]	.45*	.56 [^]	.45*	.61 [^]
33	.66 [^]	.65 [^]	.45*	.42*	.54*	.70 [^]	.71 [^]	.64 [^]	.36	.67 [^]	.69 [^]	.69 [^]	.60 [^]	.66 [^]	.66 [^]	.68 [^]	.58 [^]	.72 [^]	.47*	.57 [^]	.43*	.74 [^]	.64 [^]	.53 [^]
34	.66 [^]	.77 [^]	.51 [^]	.68 [^]	.75 [^]	.59 [^]	.58 [^]	.40*	.71 [^]	.70 [^]	.77 [^]	.85 [^]	.77 [^]	.57 [^]	.59 [^]	.60 [^]	.39*	.41*	.25	.42*	.34	.53 [^]	.42*	.44*
PP																								
35	.37	.46*	.46*	.44*	.43*	.66 [^]	.56 [^]	.61 [^]	.42*	.55 [^]	.55 [^]	.57 [^]	.50 [^]	.65 [^]	.33	.35	.28	.49 [^]	.42*	.42*	.40*	.41*	.67 [^]	.37
36	.32	.32	.29	.30	.31	.61 [^]	.51 [^]	.43*	.30	.48 [^]	.47*	.52 [^]	.35*	.62 [^]	.39*	.33	.38*	.58 [^]	.67 [^]	.58 [^]	.57 [^]	.54 [^]	.65 [^]	.33
37	.35	.39*	.38*	.41*	.40*	.53 [^]	.52 [^]	.51 [^]	.34	.50 [^]	.50 [^]	.55 [^]	.38*	.44*	.35	.30	.30	.31	.42*	.36	.44*	.37	.49*	.39*
38	.39*	.42*	.41*	.40*	.39*	.54 [^]	.55 [^]	.50 [^]	.33	.49 [^]	.49*	.55 [^]	.38*	.43*	.35	.31	.30	.32	.43*	.36	.41*	.39*	.51 [^]	.39*
ES																								
39	.39*	.16	.34	.40*	.37	.34	.32	.18	.41*	.14	.23	.32	.20	.27	-.13	-.05	.21	.11	.17	-.05	.44*	.06	.34	.36
40	.17	.23	.22	.35	.29	.35	.19	.17	.57 [^]	.24	.26	.32	.38*	.39*	.05	.11	.22	.14	.20	.27	.35	.18	.40*	.15
41	.41*	.28	.34	.35	.38*	.63 [^]	.42*	.44*	.31	.28	.32	.32	.26	.54 [^]	.36	.34	.47 [^]	.38*	.42 [^]	.58 [^]	.63 [^]	.59 [^]	.58 [^]	.46*
IN																								
42	.54 [^]	.44*	.24	.17	.29	.64 [^]	.35	.46*	.24	.47*	.51 [^]	.48 [^]	.50 [^]	.56 [^]	.18	.24	.20	.43*	.14	.44*	.53 [^]	.62 [^]	.69 [^]	.48 [^]
43	.45*	.37	.10	.20	.25	.82 [^]	.38*	.29	.29	.61 [^]	.49*	.48*	.50 [^]	.56 [^]	.37*	.34	.33	.47*	.35	.39*	.54 [^]	.68 [^]	.52 [^]	.42*
44	.61 [^]	.67 [^]	.45*	.43*	.56 [^]	.63 [^]	.65 [^]	.49*	.36	.67 [^]	.58 [^]	.72 [^]	.58 [^]	.51 [^]	.47*	.48 [^]	.21	.44*	.21	.26	.34	.51 [^]	.55 [^]	.43*
EU																								
45	.51 [^]	.34	.24	.17	.29	.63 [^]	.48 [^]	.41*	.26	.36	.34	.47*	.33	.72 [^]	.13	.18	.21	.44*	.45*	.33	.56 [^]	.47*	.70 [^]	.53 [^]
46	.56 [^]	.50 [^]	.19	.21	.31	.73 [^]	.39*	.46*	.37	.49 [^]	.45*	.49*	.52 [^]	.83 [^]	.43*	.46*	.34	.45*	.42*	.49 [^]	.37	.58 [^]	.54 [^]	.58 [^]
47	.57 [^]	.38*	.29	.25	.32	.64 [^]	.40*	.52 [^]	.38*	.24	.42*	.36	.39*	.78 [^]	.24	.26	.35	.40*	.37	.47*	.49 [^]	.48 [^]	.55 [^]	.72 [^]

Table 13 (Continued) *Intercorrelations of the ORECD Checklist Items*

Item	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	
PT																								
24																								
25	---																							
26	.85 [^]	---																						
DE																								
27	.27	.20	---																					
28	.15	.28	.64 [^]	---																				
29	.31	.32	.05	.26	---																			
CM																								
30	.64 [^]	.66 [^]	.07	.32	.14	---																		
31	.62 [^]	.69 [^]	.12	.39 [*]	.15	.88 [^]	---																	
32	.56 [^]	.60 [^]	.02	.18	.24	.75 [^]	.68 [^]	---																
33	.54 [^]	.62 [^]	.04	.48 [*]	.34	.73 [^]	.68 [^]	.63 [^]	---															
34	.41 [*]	.41 [*]	.38 [*]	.53 [^]	.34	.67 [^]	.57 [^]	.45 [*]	.62 [^]	---														
PP																								
35	.36	.47 [*]	.24	.46 [*]	.55 [^]	.36	.43 [*]	.25	.61 [^]	.44 [*]	---													
36	.31	.34	.24	.43 [*]	.68 [^]	.16	.22	.34	.58 [^]	.35	.74 [^]	---												
37	.38 [*]	.42 [*]	.12	.43 [*]	.69 [^]	.24	.27	.09	.58 [^]	.40 [*]	.73 [^]	.73 [^]	---											
38	.37	.42 [*]	.12	.43 [*]	.68 [^]	.26	.28	.09	.59 [^]	.39 [*]	.73 [^]	.72 [^]	.99 [^]	---										
ES																								
39	.44 [*]	.29	.57 [^]	.22	.17	.07	-.03	.10	.12	.26	.30	.28	.23	.23	---									
40	.13	.13	.54 [^]	.35	.22	.12	.12	.20	.04	.51 [^]	.39 [*]	.34	.04	.05	.47 [*]	---								
41	.47 [*]	.45 [*]	-.04	.05	.48 [*]	.36	.33	.56 [^]	.53 [^]	.33	.48 [*]	.66 [^]	.50 [^]	.49 [^]	.32	.36	---							
IN																								
42	.46 [*]	.43 [*]	.04	.26	.27	.40 [*]	.46 [*]	.44 [*]	.50 [^]	.36	.42 [*]	.42 [*]	.38 [*]	.37	.32	.22	.51 [^]	---						
43	.28	.37	.05	.39 [*]	.42 [*]	.35	.34	.44 [*]	.56 [^]	.44 [*]	.49 [^]	.50 [^]	.50 [^]	.48 [^]	.22	.16	.48 [*]	.74 [^]	---					
44	.43 [*]	.43 [*]	.05	.46 [*]	.27	.68 [^]	.58 [^]	.44 [*]	.74 [^]	.57 [^]	.56 [^]	.44 [*]	.65 [^]	.65 [^]	.22	.02	.41 [*]	.63 [^]	.66 [^]	---				
EU																								
45	.54 [^]	.51 [^]	.19	.24	.32	.40 [*]	.31	.60 [^]	.55 [^]	.33	.44 [*]	.49 [^]	.42 [*]	.40 [*]	.56 [^]	.22	.61 [^]	.70 [^]	.68 [^]	.58 [^]	---			
46	.51 [^]	.53 [^]	.10	.26	.41 [*]	.52 [^]	.43 [*]	.73 [^]	.60 [^]	.49 [^]	.43 [*]	.39 [*]	.30	.28	.27	.29	.53 [^]	.61 [^]	.68 [^]	.51 [^]	.81 [^]	---		
47	.72 [^]	.69 [^]	.14	.11	.36	.54 [^]	.49 [*]	.75 [^]	.52 [^]	.42 [*]	.35	.33	.25	.24	.43 [*]	.29	.59 [^]	.63 [^]	.58 [^]	.41 [*]	.84 [^]	.89 [^]	---	

Note. N=28. OE = Organizational Environment; OL = Organizational Leadership Support; KS = Knowledge/Skills Development; RE = Resources; PT = Program Theory; DE = Demand for Evaluation; CM = Communication; PP = Policies and Procedures; ES = External Support; IN = Incentives; EU = Evaluation Use.

* $p < .05$. [^] $p < .01$.

The inter-item correlations between all the items on the ORECD checklist (see Table 13) ranged from a minimum of -.01 to a maximum of .99. These correlations were mostly positive, and more than half were significant displaying moderate or high associations ($r=.38$ to $.99$). The highest correlation ($r=.85$) found within items from different components was between Item 12 (Organizational and Leadership Support component) and Item 34 (Communication component), followed by the correlation ($r=.83$) between Item 14 (Organizational and Leadership Support component) and Item 46 (Evaluation Use component). Very few negative inter-item correlations ($r= -.01$ to $-.13$) were found. The magnitude of almost all of them was close to zero and none of them were significant. These negative correlations are the lowest correlations yielded between all the items.

Appropriateness of the ORECD Checklist for the Intended Uses

The questions in the Questions for Field Study form asked nonprofit organizations to provide their opinion about the ORECD checklist in three main areas: (a) usefulness of the checklist to determine the readiness of the organization to develop evaluation capacity; (b) helpfulness of the checklist to guide the organization in getting ready for developing evaluation capacity; and (c) necessary changes to improve the checklist. A set of codes were developed by the researcher based on the questions and participants' responses. The data were coded according to the following codes: *useful*, *somewhat useful*, *not at all useful*, *create awareness*, *very helpful*, *somewhat helpful*, *depend on the situation*, *checklist length*, *yes*, *mostly*, *somewhat*, *not sure*, *no*, and *modify wording*.

Information from participants about the usefulness of the ORECD checklist to determine the readiness of the organization to develop evaluation capacity was coded as *useful* or *create awareness*, more than any other code. Responses coded as *useful* were those in which

participants expressed how the ORECD checklist was beneficial for the organization to develop evaluation capacity. Some examples are, “The checklist will be a great tool in helping the organization to develop an evaluation development plan,” “Good; questions were relevant,” “Very useful,” “The checklist is very helpful in providing the appropriate categories for assessment of any organization,” and “Covered broad range very well.” Responses coded as *create awareness* were those that illustrated how the ORECD checklist informed participants about the needs of the organization. For instance, “It was a good tool to point out where we were lacking and where our strengths are,” “Got us thinking about planning and needs we have not thought about before,” “Very helpful to identify areas that need strengthening,” and “The checklist reveal our organization lack of readiness.”

There were few instances in which participants indicated that the ORECD checklist was *somewhat useful* for the organization and so were coded. Responses include, “Somewhat useful; it doesn’t seem to allow for small organizations that understand the value of evaluation but lack sufficient resources to support it,” “Seemed a bit repetitive,” and “I’d probably feel most comfortable in speaking with someone with expertise in the field to determine how or if we should devote additional resources to this subject.” Also, one participant explained that the ORECD checklist was *not at all useful* because “Our organization is small; four volunteers officers without evaluations completed.”

Feedback about the helpfulness of the ORECD checklist to guide the organization in getting ready for developing evaluation capacity was coded as *very helpful* when responses were completely positive about it. The majority of the responses to this question were classified within this code. Examples consist of, “Excellent guide that can be used throughout the organization and with the board,” “It is helpful in our efforts to take the first step (discussion)

toward the decision to form and evaluation plan,” “It provides a step-by-step rubric. Therefore, great directions are given to the process of guidance,” “It is a good starting point,” and “It was very helpful because it addressed many areas, levels, and components of organizational evaluation.” Responses were coded as *somewhat helpful* when participants expressed that the checklist is helpful to some extent and when they will need outside support to use the ORECD checklist. Their responses were, “It will be somewhat helpful because it breaks down the concept of evaluation into actionable categories,” “It would require significant outside assistance and training. I view it as a valid process but limited use due to our small size,” and “I’d likely seek out someone with evaluation experience to help guide my thinking rather than using the checklist.”

There were also responses coded as *depend on the situation* when participants indicated that the utility of the checklist can vary according to the situation in the organization. Examples are, “Could give insight when establishing an evaluation procedure,” “I need to get the board more involved,” “[It is helpful] only if we have a regular turn out of volunteers,” and “Very helpful only if such a plan is accepted and would benefit the organization, which I think it would.”

Participants’ responses about necessary changes to improve the ORECD checklist were for the most part, straightforward (e.g., yes, no, not sure) and positive. Responses coded as *yes* were those in which participants indicated that the items were clear, the response scale was appropriate, and the length of the ORECD checklist was right. Almost all of them responded only with a “yes”. Additional examples are, “It seems the right length” and “just right [length of the checklist].” Responses coded as *mostly* were those in which participants expressed that, in general, the items were clear and the response options were appropriate. The majority of them

responded saying “mostly.” An additional example is, “Some questions were very long, but examples were good.” There were few instances in which participants indicated that the items were somewhat clear and/or the response options were somewhat appropriate but their feedback did not provide specific recommendations. For example, “There are nuances to a majority of the items that can’t easily be represented on a scale such as this.”

Very few responses were coded as *modify wording*. For instance, “Some of the checklist language presumes a greater knowledge of ECD than might be the case” and “The section for evaluation use seems to jump from evaluating readiness to assuming adoption.” Responses were coded as *checklist length* when participants expressed that the ORECD checklist was either long or short. Some of the responses are, “Too long, appears too repetitive” and “A little lengthy but covered a lot.” Another response was, “Appropriate but it could have been longer to ensure a thorough investigation.”

Responses coded as *no* are those in which participants expressed that it is not necessary to add items to or delete items from the ORECD checklist. The majority of the participants responded just saying “no.” Additional examples are, “None come to mind,” “None I am aware of,” and “Can’t think of any right now.” Responses coded as *not sure* indicated the possibility of adding items to the ORECD checklist. Examples include, “Not sure, can use discussion” and “Perhaps questions pertaining to awareness of types of evaluation, access to resources, and understanding of evaluation methods.” Overall, participant responses indicated that the ORECD checklist is appropriate for the intended uses.

Potential Consequences of Using the ORECD Checklist

The five executive directors interviewed provided information about potential consequences related to the use of the ORECD checklist. Specifically, the interview examined

information about the process these participants followed during the completion of the ORECD checklist and benefits or issues they foresee as a result of using this tool. It also explored recommendations for improvement in order to complement the feedback provided by all the nonprofit organization participants in the field study. Overall, the data from the interviews were coded according to the following codes: *action taken*, *checklist content*, *benefit*, *valuable aspect*, *no issues*, *time resource*, and *generate interest*.

According to the executive directors, the amount of time necessary to complete the ORECD checklist ranged from 30 to 45 minutes. They also commented about the physical format of the ORECD checklist indicating that it was helpful, understandable, and easy to follow. In addition, participants found the response options appropriate.

Participants' responses about the process they followed to complete the ORECD checklist were coded either as *action taken* or *checklist content*. Responses coded as *action taken* were those in which participants provided information about the sources they used to get the information they needed to complete the ORECD checklist. The responses also included whether they completed the ORECD checklist by themselves or with assistance from someone else to come up with the best responses about their particular situation. They indicated, "I based it upon as the position that I am in as executive director and what I had then...and what I was finding from not only the board of directors but also the one staff member I had..." "I didn't consult with anybody else, not at this time," "I had to take some time to work through it. It was self-explanatory, I just went question by question," "There is no one that I needed to turn to other than may be my accountant for any information. This was all me," "I just took from my knowledge and filled it out. There was no checking with anybody else," "I don't think we did it in one time because I gave it quite a bit of thought. I didn't rush through it," and "We read the

questions and thought about it...and we circle the number. There were couples of questions that I discussed with the board. I took them into account.”

Responses coded as *checklist content* were those in which participants commented about how the content of the ORECD checklist relates to what the organization is doing, the size of the organization, and when they expressed any type of situation completing the ORECD checklist. For example, “...it deals with a lot of things that I was already doing. It is more of what I had happening,” “...thinking about a few things more in depth...what is happening in our organization and how these relate with the criteria up here,” “The only thing I found difficult at different times was, there is always a question that is like another question and then is like what else are you meaning with this?...I don’t remember any of those questions,” “My organization is small. We are in the process of trying to do evaluation, so there wasn’t much that I could relate to,” and “...there were a couple of questions that I wasn’t sure if I was answering them the right way. I just went with what was my closest idea. There wasn’t very difficult; it was a fluid process.”

Potential benefits for the organization as a result of using the ORECD checklist were coded as *benefit* and *valuable aspect*. Responses coded as *benefit* were those in which participants expressed that the ORECD checklist produced or will produce positive effects. In this regard, they indicated that the ORECD checklist was or will be helpful for the organization for a variety of reasons. Specifically, it provides guidance about what the organization needs to develop evaluation capacity, facilitates determining what the organization already has in place in order to develop evaluation capacity, and helps improving evaluation in the organization by using the ORECD checklist as a planning tool. Examples of their responses are, “It was very helpful because it allowed me to see what we are not doing and as a planning tool to start

implementing this,” “It’s a catalyst, it is a wake-up call...I have to start thinking about evaluating and measuring, and those metrics are going to be very important in the future for getting grants,” “It was helpful for picking out things that I needed to capture as I was going forward on what I was doing,” “I went ahead and started the process of implementing what I need to do with it...it was helpful on one part of it and the other part is already in place,” “...it made me think about a lot of different components that I was aware but it was articulated to the question. Once I read the question I said, oh yeah, I do need that,” “I thought it will give us the opportunity to look at some new ways to evaluate what we are doing and possibly make changes and fix some things,” and “...we have started talking about some of the things that we do in a different way...and some of your language [in the checklist], we have internalized that....Everybody is kind of inquisitive just by having reviewed the language”

There was only one participant who expressed uncertainty about the benefits of using the ORECD checklist as a result of what was taking place in the organization, but expressed interest on keeping a copy of the ORECD checklist so in the future they know what they need to do. This participant said, “Not sure at this time [about benefits] because we are not ready to tackle the project...It probably will be [beneficial] when tackle this on a grander scale.”

Participants’ responses coded as *valuable aspect* where those in which participants highlighted something about the ORECD checklist that they found useful for the organization. Examples are, “There is a couple of items in there that really made me relook to a couple of things that I was needing to do, [that] I haven’t thought of it,” “...these are very important questions if we want to move forward,” “...there were some [aspects] more beneficial than others...priority things will be budget, the financial focus, the time resource...the training of the staff is another thing that is important because I want to have quality performance....Also,

communication it's a big piece," "I thought your questions were interesting and I thought they were applicable," and "What you have asked was very specific in a lot of different ways."

Similarly, participants offered feedback about potential issues with the ORECD checklist. Most participants reported that they do not foresee any issues or problems for their organizations as a result of using the ORECD checklist. Accordingly, the majority of the responses were coded as *no issues*. Examples responses are, "No, I don't see any problems" and "No. First I wasn't sure if it applies to us but then as I considered the questions it was like, oh yeah, this would apply to anybody." The only concern raised by one participant was coded as *time resource* as it relates to the time necessary to implement a plan to develop evaluation capacity. This participant explained that, "The only issue will be the time it will take to implement a plan taking away from the work that we need to do, because most of my staff is part-time and it is all program oriented."

Participants also talked about their experience using the ORECD checklist. The responses were coded as *generate interest* and consisted of participants' comments about the ORECD checklist as a mechanism to move toward evaluation. They commented that, "This is a perfect way to get the board of directors more involved; for the most part they are just names right now. When you are [a] small [organization] they are your staff," "We have talked about what is evaluation; self-evaluation; and evaluation of our processes...we want to be self-evaluated, as well as across the board, [to see] what we are doing in our work," "I know this [evaluation] is important," and "I think it will be a good tool to train executive directors on what all the different things that are needed to think about when you do, when you pursue evaluation."

When asked about further recommendations for improvement of the ORECD checklist, participants said that they did not have any additional suggestions. In fact, some of them used

this opportunity to express that the ORECD checklist was very good or well done. One commented that was glad to be part of the study, indicating about the ORECD checklist, “is not anything that I don’t know already but formalizing it, make it serious, and getting a copy of it, will be very important to me.” Overall, participant comments showed for the most part, positive responses, indicating that the ORECD seems appropriate and beneficial for the intended uses.

Final Adjustments to the ORECD Checklist

Decisions about the ORECD checklist were made taking into consideration the results from the analysis of the psychometric properties and the feedback from participants. These were about retaining items without changes, modifying the wording of the items, moving items to another component, merging, and adding items. As in previous phases, the action taken with each individual item is presented in Table 14. It is important to note that as a result of the field study changes to the introduction and directions of the ORECD checklist were not required, and all the components were retained. Regarding the rating scale, the names of the response options were retained but the numerical values were removed. Estimating a total value for the overall ORECD checklist or by component will not provide organizations the information necessary to determine their strengths and the specific areas that need to be prioritized. Thus, it will be necessary to examine the items individually giving preference to those marked as *to a small extent* and *not at all* in order to support the development of evaluation capacity.

An exploration of the content of Item 29 and Item 41 was necessary. It was found that these items decrease the alpha of their components to less than .70. The content of Item 29 represents an important piece of ECD. As a result, it was reworded and moved to the Organizational Leadership Support component which seemed to be a better fit. Therefore, a new item was added to Demand for Evaluation because typically, a minimum of three items is

recommended to provide more information (Hatcher, 1994). The purpose at this point is to have enough items to explore through factor analysis. Regarding Item 41, although the item-to-total correlation is greater than .3, it decreases the alpha of the component to less than .70. Thus, the item was reworded to make it clearer and align it better to the content of the component.

Table 14

Action Taken with the Items of the ORECD Checklist After the Field Study

Original Item	Action	New Item
1. The internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).	R	
2. Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).	R	
3. There is a general understanding of how evaluation can provide important information to the organization.	R	
4. There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.	R	
5. There is a general commitment to learning from evaluation (process and/or results) throughout the organization.	R	
6. The organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.	R	
7. Staff is aware of how their work relates to evaluation.	R	
8. There is organizational stability (for example, the organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and the staff turnover is low).	M	There is organizational stability (some aspects to consider are: the organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, has control of the finances, and the staff turnover is low).
9. The organizational leadership supports the development of internal evaluation capacity.	R	

Table 14 (Continued) *Action Taken with the Items of the ORECD Checklist After the Field Study*

Original Item	Action	New Item
10. The organizational leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).	R	
11. The organizational leadership is committed to encouraging evaluation capacity development activities.	R	
12. The organizational leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.	R	
13. The organizational leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).	R	
14. The organizational leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.	M	The organizational leadership is committed to investing in training/professional development (including for example, types of evaluation approaches and methods of data collection) to increase the evaluation skills/knowledge of the staff involved in ECD.
15. There is a plan to identify the strengths/weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).	R	
16. There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work vs. traditional lectures).	R	
17. There is staff with evaluation expertise/ experience conducting evaluations inside the organization in order to support the ECD process.	M	There is staff inside the organization with evaluation expertise/experience conducting evaluations in order to support the ECD process.
18. There are strategies in place to be able to access sufficient human resources for the ECD process.	R	
19. Organization has committed financial resources to develop evaluation capacity.	M	Organization can commit financial resources to develop evaluation capacity.

Table 14 (Continued) *Action Taken with the Items of the ORECD Checklist After the Field Study*

Original Item	Action	New Item
20. There is a plan for securing additional fiscal resources to develop evaluation capacity.	M	There is a plan for securing fiscal resources to develop evaluation capacity.
21. An evaluation budget is a priority for the organization.	R	
22. There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.	R	
23. There is a plan to provide staff sufficient time during the workday to work on evaluation activities.	M	There is a plan to provide staff sufficient time during the workday to work on ECD/evaluation activities.
24. The goals/objectives of the organizational program(s) are well defined.	R	
25. Staff has a common understanding about what organizational program(s) do.	R	
26. Staff has a common understanding of how organizational program(s) work (for example, there is a logic model or other graphical representation of the program(s) and the expected outcomes).	R	
27. There is demand for evaluative information from external stakeholders, such as funders (for example, on what works, accountability requirements, and strategies to increase evaluation use).	R	
28. There is demand for evaluative information from internal sources.	R	There is demand for evaluative information from internal sources (for example, to increase funding sources, for program improvement).
29. There is a written plan about how to develop evaluation capacity.	R	The organizational leadership has a written plan about how to develop evaluation capacity.
30. There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).	R	
31. There are communication structures to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	R	
32. There are communication procedures to manage/share information (such as, collection, dissemination, and disclosure of information).	R	

Table 14 (Continued) *Action Taken with the Items of the ORECD Checklist After the Field Study*

Original Item	Action	New Item
33. There is a feedback mechanism (action or means used to modify the ECD process as a result of information received) to facilitate learning within the evaluation capacity development process.	M	There is a feedback mechanism (action or means used to modify a process as a result of information received) to facilitate learning within the evaluation capacity development process.
34. The organizational leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.	R	
35. There is a plan in place to develop ECD policies (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).	M	There is a plan in place to develop evaluation policies (including for example, rules to guide evaluation decisions, actions and activities; evaluation training for staff; ethical considerations; and disclosure of information).
36. There is a plan in place to develop ECD procedures (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).	M	There is a plan in place to develop evaluation procedures (these are step by step instructions to put policies into action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in any ECD and evaluation consultation processes).
37. There is a plan to make ECD policies accessible to all staff.	M	There is a plan to make evaluation policies/procedures accessible to all staff.
38. There is a plan to make ECD procedures accessible to all staff.	D	
39. External stakeholders (for example, funders) support the development of evaluation capacity.	R	
40. Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.		
41. There is a plan to establish mechanisms for advancing the development of evaluation capacity (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).	M	There is a written plan to establish mechanisms for advancing the development of evaluation capacity by joining efforts with external sources (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).
42. Incentives are available to encourage staff participation in the development of evaluation capacity (for example, allotted time and flexibility for people to incorporate evaluation into the everyday work of the organization).	R	

Table 14 (Continued) *Action Taken with the Items of the ORECD Checklist After the Field Study*

Original Item	Action	New Item
43. There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.	R	
44. Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).	R	
45. The organizational leadership promotes the dissemination of evaluation results (to inform staff about them as appropriate).	M	The organizational leadership has identified means to promote the dissemination of evaluation results as appropriate (for example, meetings, reports, newsletters).
46. The organizational leadership promotes the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).	M	The organizational leadership has identified means to promote the use of evaluative information (for example, giving access to the information, translating the results into appropriate action, and exploring the positive and negative implications of using the results).
47. The organizational leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).	M	The organizational leadership/staff is willing to make use of evaluation results (for example, for planning, decision-making, deciding how to implement, deliver, and improve programs, and identifying what has been effective).

Note. D = dropped; M = modified to add content or reword it; R = retained without changes

Other modifications included the improvement of the example in Item 8 and the addition of an example to Item 14, to incorporate recommendations from participants, and to Item 28, to facilitate its understanding. Also, Item 17 was reworded to improve clarity of the content, with the anticipation that this will improve even more the alignment of the item (i.e., item-to-total correlation) to the component. Item 19 and Item 20 were somewhat reworded to better denote organizational readiness. In addition, information was added to Item 23 to include the time necessary during the workday to work on ECD. Moreover, the information inside the parenthesis in Item 33 was slightly reworded to make it more inclusive and applicable to the

various processes that can occur in any organization, with the potential to facilitate learning within the ECD process.

After examining Item 35 and Item 36, and taking into consideration a remark from one of the experts from Phase 2, the content of these two items was modified. The purpose was to improve how these items portray what usually happens during the development of evaluation capacity. The content of these items emphasized the development of ECD policies and procedures, respectively. However, what is really essential is to plan the development of evaluation policies and procedures to promote the institutionalization of evaluation in an organization, which is the ultimate goal of ECD. Further modifications included Item 37 and Item 38. These two items were combined because the correlation yielded between them was almost perfect, suggesting that they were measuring the same idea. The revisions to the ORECD checklist also included Item 45, item, 46, and Item 47. These were reworded to better represent readiness and the examples were improved accordingly.

As a result of the revisions made to the ORECD checklist after the field study, the final version produced in Phase 4 consisted of 47 items across 11 components. These components and the amount of items in each one are: (a) Organizational Environment, 8 items; (b) Organizational Leadership Support, 7 items; (c) Knowledge/Skills Development, 3 items; (d) Resources, 6 items; (e) Program Theory, 3 items; (f) Demand for Evaluation, 3 items; (g) Communication, 5 items; (h) Policies and Procedures, 3 items; (i) External Support, 3 items; (j) Incentives, 3 items; and (k) Evaluation Use, 3 items.

Summary of Results by Research Question

1. To what extent does the Organizational Readiness for Evaluation Capacity Development Checklist integrate elements presented in the literature that support evaluation capacity development/building based on expert review?

The review of relevant literature presented provided content validity evidence. In addition, the review of the content by experts provided face and content validity evidences of the ORECD checklist. Even though results from the experts' feedback were mixed, most of the items with the largest discrepancies among experts were removed. Also, experts offered varied recommendations which explain why the agreement among them was estimated as fair. However, they indicated that the ORECD checklist seems to be a good measure of the elements necessary for evaluation capacity development.

2. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist suitable for the intended users as perceived by individuals with formal evaluation training?

The review by participants with formal evaluation training provided predominantly positive feedback about the appropriateness of the wording and format of the ORECD checklist. The majority of the recommendations for improvement consisted of minor grammatical or structural changes of the items and there was no need to remove any item.

3. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist appropriate for the intended uses as perceived by nonprofit organizations?

Participants in the field study indicated that the ORECD checklist will be useful to determine the readiness of the organization to develop evaluation capacity. They believe that the ORECD checklist creates awareness about the needs of the organization and will be very helpful guiding

the organization in getting ready to develop evaluation capacity. Very few responses suggested the need of modifications to the ORECD checklist and some were adopted, as applicable.

4. What are the potential positive and negative consequences of the Organization Readiness for Evaluation Capacity Development Checklist?

The favorable responses from participants indicated that the consequences of using the ORECD checklist will be mostly positive. They believe that the ORECD checklist facilitates determining what the organization has already in place to develop evaluation capacity and it can be used as a planning tool and mechanism to move toward evaluation. In addition, most participants do not foresee any issues as a result of using the ORECD checklist.

5. What are the psychometric properties of the Organizational Readiness for Evaluation Capacity Development Checklist in the field study?

The internal reliability for most of the components was good or excellent, with the exception of two components that showed questionable values. Almost all of the items on the ORECD checklist exhibited moderate to high item-to-total correlation. Also, all the correlations between components were positive, most of them were significant, and more than half showed high associations. Similarly, all the inter-item correlations by component were positive and for the most part significant, displaying the majority of them moderate and high associations. In addition, the inter-item correlations between all the items on the ORECD checklist revealed that most of them were positive and more than half were significant, showing moderate and high associations.

CHAPTER V

DISCUSSION

This chapter presents a summary and discussion of the findings as well as conclusions drawn about the validation process of the Organizational Readiness for Evaluation Capacity Development (ORECD) Checklist. It also presents the limitations of the study, recommendations for future research, including the need to obtain additional validity evidence of the ORECD checklist, and implications for practice.

Summary of the Study

Evaluation capacity development has been acknowledged as a system of processes to help organizations achieve sustainable evaluation practice, but its advancement varies depending on different factors and the situation in each organization, considering that organizations are changing entities. Therefore, it is necessary to examine the existing evaluation capacity and the status of that system in the organization before starting an ECD process. In this way, there will be increased possibilities of success, determined by the establishment or strengthening of an evaluation system into the organization.

Although, there are other checklists that address the ECB/ECD area, none of these checklists were designed to evaluate the readiness of an organization to embark in the development of evaluation capacity at any given point. To respond to this need, the ORECD checklist was designed to guide stakeholders in determining the readiness of their organization to develop or strengthen its evaluation capacity. It is important to note that some ECD areas are

going to be more critical than others. Some of these areas include organizational leadership support and resources, particularly; human and financial resources, appropriate technologies, and time will be a priority.

As part of the study, the ORECD checklist was designed during Phase 1 based on a review of relevant literature in areas such as ECD/ECB and internal evaluation in organizations with emphasis in nonprofit organizations. After designing the ORECD checklist, a validation process followed to generate a body of evidence and determine its strengths and limitations, using both quantitative and qualitative procedures. Evidences, including face, content, and consequential validity were gathered from different sources. Specifically, Phase 2 consisted of a review of the ORECD checklist by relevant experts in order to obtain face and content validity evidences. Feedback about different aspects was obtained, including relevance and clarity of the items, the extent to which the items fit each of the components, and comprehensiveness of the ORECD checklist to represent the construct. Phase 3 consisted of a pretesting to determine the appropriateness of the wording of the items (clarity) and format of the ORECD checklist, according to doctoral graduate students with formal training in evaluation and professional evaluators. Lastly, Phase 4 consisted of a field study in which nonprofit organizations used the ORECD checklist and provided feedback about it. During this phase, additional feedback was obtained through face-to-face interviews with some of the participants to complement the information received. Moreover, this phase produced data about the psychometric properties and also provided information about the utility and benefits of the ORECD checklist, in order to establish consequential validity. Overall, Phases 2 to 4 involved a comprehensive review of the ORECD checklist according to the feedback received which is reflected in the modifications and refinement of the several versions of the ORECD checklist generated.

Discussion of Findings

The discussion of the findings follows and it is presented in the order of the research questions that guided this study. The research questions addressed were:

1. To what extent does the Organizational Readiness for Evaluation Capacity Development Checklist integrate elements presented in the literature that support evaluation capacity development/building based on expert review?
 - a. What is the evidence for face validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
 - b. What is the evidence for content validity of the Organizational Readiness for Evaluation Capacity Development Checklist?
2. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist suitable for the intended users as perceived by individuals with formal evaluation training?
3. To what extent is the Organizational Readiness for Evaluation Capacity Development Checklist appropriate for the intended uses as perceived by nonprofit organizations?
4. What are the potential positive and negative consequences of the Organization Readiness for Evaluation Capacity Development Checklist?
5. What are the psychometric properties of the Organizational Readiness for Evaluation Capacity Development Checklist in the field study?
 - a. What are the item-to-total correlations for each component?
 - b. What is the inter-item relationship for each component?
 - c. What is the relationship between the components of the checklist?
 - d. What is the relationship between all the items in the checklist?

- e. What is the internal consistency for each component?

Face and Content Validity Evidences of the ORECD Checklist

In the Standards (AERA, APA, & NCME, 1999) it is explained that different sources of evidence can elucidate and contribute to validity, supporting the proposed interpretation of the results of an instrument. Following this, face, content, and consequential validity evidences were obtained as part of the validation process of the ORECD checklist. Face and content validity evidences were collected during the initial phases of the study and consequential validity was addressed during the last phase of the study. A discussion of face and content validity evidences follows. Consequential validity evidence is discussed later on in this chapter.

Face validity evidence was established by experts who indicated that the ORECD checklist seem to be a good measure of the elements necessary to determine the readiness of an organization to develop evaluation capacity. Content validity evidence for the ORECD checklist was obtained from two different sources. One way in which content validity can be established is through a logical analysis of the relevance of an instrument to represent the content domain and support the interpretations of the results (AERA, APA, & NCME, 1999). Therefore, the review of the literature was utilized to obtain content validity evidence, identifying and examining the adequacy of the content of the items and the components of the ORECD checklist according to the characteristics and description of the content domain found in the literature. This evidence was obtained by the researcher.

Similarly, content validity evidence can come also from experts in charge of judging how the content of the instrument is related to the construct and also determining any potential difficulty with the items due to the wording used (AERA, APA, & NCME, 1999). The agreement among these individual experts determines the representativeness of the content

domain by the set of items developed and also provides information about the clarity of the items. In this study, agreement among experts was obtained through qualitative procedures and also it was estimated using the Intraclass Correlation Coefficient (ICC). Differences among experts were evident in the variety of recommendations provided by them to improve some of the items. Also, most of the ICC values obtained for both relevance and clarity of the items were considered fair. These values represent the differences in the ratings awarded to the items. However, an examination of the individual ratings assigned by experts to each item revealed that most of the time, items were rated as *relevant* or *very relevant*. Only one expert rated 14 items as *not relevant*. Ten of these items were removed from the ORECD checklist and the other four items were retained because the other experts rated them mostly as *relevant* or *very relevant*. In addition, an exploration of the individual ratings assigned by experts to each item concerning clarity revealed that over half of the time experts rated them as *very clear* or *somewhat clear*, *need revision*. As a result, all the items with a mean rate less than 3.00 for relevance were removed and most of the retained items were reworded or modified in order to address the issues highlighted by experts and support content validity.

The ECB/ECD literature provides some explanations that help to understand the partial consistency among experts regarding the relevance of the content. The aspects found on the literature that might contribute to the results of this study are: (a) ECD/ECB is a relatively recent conceptual development or area that is still being defined (King & Volkov, 2005; Labin et al., 2012; Nielsen, Lemire, & Skov, 2011); (b) there are several ECB/ECD definitions, something expected and encouraged to promote the development of new ideas (Compton & Baizerman, 2007); (c) it is complex to define an idea like ECB/ECD because it is highly context dependent, making its conceptualization diverse (Taut, 2007); (d) ECB/ECD has been recognized as a

multidimensional construct (Kirsh et al., 2005, p. 235); (e) the ECB/ECD process is not completely specific and includes somewhat unclear practices and little systematic assessment (Baizerman et al., 2002); and (f) “the evaluation profession as a whole still lacks a well-developed theory and associated indicators for understanding evaluation capacity at an organizational level...” (Milstein et al., 2002, p. 41).

In fact, in an effort to describe the ECB/ECD construct, several conceptual models and frameworks have been developed, however, “slight attention has been given to the empirical validation of such models” (Suárez-Balcázar & Taylor-Ritzler, 2014, p. 96). Also, the focus or emphasis of these models differs, showing the existent difficulties to clearly delineate the ECD concept. Therefore, as ECB/ECD matures, more research and attention is needed to better comprehend what it takes to successfully sustain this process and gain understanding of its practices, so they can inform one another (Huffman et al., 2008; Stockdill et al., 2002; Suárez-Balcázar & Taylor-Ritzler, 2014; Taylor-Powell & Boyd, 2008).

Suitability of the ORECD Checklist for the Intended Users

The same way it is necessary to make certain that the content of the items on an instrument is an appropriate representation of the content domain, it is also important to ensure that the wording of the items is suitable for the intended users. Items that are difficult to understand, have inappropriate wording or unnecessary jargon should be examined (AERA, APA, & NCME, 1999). As part of this study, the clarity of the items was evaluated by experts during Phase 2. Consequently, the wording of many items was modified according to the feedback received to overcome any validity issue. After modifications in Phase 2, the clarity and readability of the items were examined again by doctoral graduate students with formal training in evaluation and professional evaluators in Phase 3, in order to capture any additional problems.

The aim of this second examination was to determine if the wording of the items seemed appropriate. Participants in Phase 3 believed that in general, the wording of the items on the ORECD checklist was appropriate. The majority of the recommendations for improvement consisted of minor grammatical or structural changes of the items and none of the participants stated the need of removing items from the ORECD checklist.

In addition to the wording of the items, the response scale and format (flow of the items and physical format) of the ORECD checklist were also reviewed. These were not modified because the feedback received from participants was mostly positive or not applicable, considering the purpose for which the ORECD checklist was designed. Overall, they considered the ORECD checklist suitable for the intended users.

The clarity of the items was also evaluated in Phase 4. Participants (i.e., intended users) had the opportunity to provide feedback about the items. Most of the responses were favorable indicating that the items were clear. Few participants provided specific recommendations and these were taken into consideration when making the final modifications to the items. Predominantly positive responses suggested that the wording of the items is appropriate for the intended users.

Consequential Validity Evidence of the ORECD Checklist

The final phase of this study addressed consequential validity. The aim of this type of evidence is to obtain information about the implementation and use of an instrument and the decisions made based on the interpretations. The purpose of collecting this type of evidence is to determine if the benefits of using the instrument are likely to be realized (AERA, APA, & NCME, 1999). Also, it is important to demonstrate that unintended consequences are minimal

(Brualdi, 1999). Accordingly, data were gathered from participants (i.e., intended users) in Phase 4 through a feedback questionnaire and face-to-face follow-up interviews.

The information collected from participants suggests that using the ORECD checklist produced or will produce positive effects and will be beneficial for organizations to determine their readiness to develop evaluation capacity. For the most part, participants found the content of the ORECD checklist relevant and very useful. They also believe that the ORECD checklist creates awareness about the strengths and the needs of the organization, specifically, assisting in the identification of the things they are lacking and areas that require strengthening. This finding concurs with part of the ECB goal which indicate that it "...create awareness and support for program evaluation and self-evaluation as a performance improvement strategy ..." (King & Volkov, 2005, p. 11). It was also found that participants' responses recognized the ORECD checklist as a helpful instrument to guide organizations in getting ready for the evaluation capacity development process, by determining what the organization has already in place and creating a plan to develop evaluation capacity. Specifically, they indicated that the ORECD checklist is a good starting point and can be used as a planning tool and mechanism to move toward evaluation.

Even though participants do not foresee any issues as a result of using the ORECD checklist, there were some concerns. One of the participants interviewed expressed uncertainty about the benefits of the ORECD checklist because the organization was not ready to embark in the development of evaluation capacity. Another participant mentioned that the amount of time necessary to implement a plan to develop evaluation capacity can be a limitation when most of the staff is part-time and designated to provide direct services. This concern is mentioned in the literature as one of the major constraints to develop and implement evaluation capacity (Arnold,

2006; King & Volkov, 2005; Kirsh, 2005). Also, one mentioned that they will need outside assistance and training to use the ORECD checklist. Carman and Fredericks (2010) present similar findings explaining that is common for nonprofit organizations reporting the need of technical assistance to improve their evaluation capacity.

There was another instance in which few participants expressed that the utility of the ORECD checklist is limited when the organization is small. This situation was more common for those small organizations that only have volunteers. However, this finding was not at all unexpected because as Huffman et al. (2008) indicates, developing evaluation capacity is challenging and it includes a broad range of difficulties and issues. Nevertheless, this concern was not shared by all the small organizations that participated in the study. Some of them were positive about the possibility of developing evaluation capacity and even one mentioned the importance of involving the board of directors as an option to make the ECD process feasible. Despite developing evaluation capacity is complex, “the capacity can be built, slowly and systematically over time, through procedures that make sense even for small organizations” (King & Volkov, 2005, p. 12). Overall, participant favorable responses showed, for the most part, positive consequences when using the ORECD checklist, providing evidence that it is appropriate and beneficial for the intended uses.

Psychometric Properties of the ORECD Checklist

Some decisions about changes to the ORECD checklist were made based on the psychometric properties. The results including Cronbach’s alpha and correlations were central to guide those decisions.

Internal consistency. Cronbach’s alpha was estimated to determine the internal consistency reliability of the ORECD checklist components. The alpha coefficients for each of

the components on the ORECD checklist showed mixed results but, for the most part, displayed excellent internal consistency reliability, suggesting that the items are measuring aspects of their corresponding component. The internal reliability for most of the components of the ORECD checklist, presented in Table 9, was excellent (.90 to .94) or good (.85 and .86), with the exception of two components, External Support and Demand for Evaluation, which showed alphas close to but below .70. These alphas were considered questionable suggesting the need of revisions.

In the Demand for Evaluation component, Item 29 displayed an item-to-total correlation of .22. This value was considered low indicating the need to revise or remove the item. Removing Item 29 from the component would substantially increase the internal consistency reliability from .61 to .77, making the alpha an acceptable value. Thus, the content of this item was examined to determine its relevancy and consider the possibility of removal. It was found that the content of the item is relevant for the readiness for evaluation capacity development construct, as rated by experts in Phase 2. This item represents the importance of having a written plan to develop evaluation capacity. One expert indicated that "...if people don't have an actual, physical plan then ECB can easily go by the wayside." As a result, the item was retained, reworded, and moved to the Organizational Leadership Support because this component seemed a better fit for this item. Retaining the item will give the opportunity in future studies of the ORECD checklist to find out how it works.

Moving Item 29 left the Demand for Evaluation component with only two items. Nevertheless, having at least three items representing a component can make it more valuable because more information is provided (Hatcher, 1994). Therefore, an item was added to this component. The item added was previously designed as part of the initial ORECD checklist and

substantially modified to represent the importance of demand for evaluation from internal sources. However, going back to the literature about the topic, it was found that the original item designed is relevant and may contribute with valuable information, in addition to the items retained in that component. Thus, the following item was incorporated to the final version of the ORECD checklist: *There is demand to develop evaluation capacity (either internal/external source)*. The need of this additional item, as well as the other items on the ORECD checklist, to represent a component will be determined in the future through factor analysis.

In the External Support component, Item 41 showed an item-to-total correlation of .36. Although this value is above the minimum recommended level of .3, if the item is removed, the alpha will increase from .66 to .70. This represents a small increase of the alpha, meaning that this item is contributing in some way with information. Therefore, an exploration of the content of this item was conducted. It was determined that the alignment between the idea this item was trying to convey and the content of the component it may not have been totally apparent for participants. As a result, the item was reworded and retained for future exploration, including a factor analysis.

The item-to-total correlation of all the other items on the ORECD checklist was also examined. These items showed moderate to high item-to-total correlations, indicating that they contribute rich information to the ORECD checklist. In addition, the value of the alpha if an item is deleted was compared to the alpha obtained for each component to identify the impact of removing each item. It was found that, despite item 17 has an item-to-total correlation of .54, if the item is removed the alpha of the component will increase from .85 to .95. Therefore, the item was carefully examined and reworded in an effort to make it an even better fit for the component.

Correlations. Correlations between components, inter-item correlations by component, and correlations between all the items on the ORECD checklist were estimated using Pearson product moment correlation coefficient. All the correlations between the components of the ORECD checklist were positive as expected, and almost all of them were significant, except two (see Table10). The correlation between Knowledge/Skills Development and External Support components ($r=.31$) and between Program Theory and Policies and Procedures components ($r=.34$), even though moderate in magnitude, they were non-significant. These values could be influenced by the sample size ($n=32$). In addition, more than half of the correlations between components were high and the remainders were moderate, indicating substantial relationship between the components of the ORECD checklist. This suggests that the components are measuring various aspects of the same construct.

Similarly, all the inter-item correlations by component were positive as expected and were mostly significant. The majority of these correlations were high, few of them were moderate, and only three of them were low indicating for the most part, substantial relationship between the items in almost all the components. This suggests that most of the items are measuring aspects of their corresponding component. Specifically, in the Organizational Environment component it was found that the correlation between Item 4 and Item 8 was low. Also, even though Item 8 had moderate correlations with Item 3, Item 5, and Item 7, these were non-significant, suggesting that Item 8 is measuring a unique aspect of this component. Another component with low correlations was Demand for Evaluation. In this component, the correlations between Item 27 and Item 29 and between Item 28 and Item 29 were low, meaning that there was a minimal relationship. As previously discussed in this chapter, this item was reworded and moved to another component. Moreover, two of the three correlations in the

External Support component even though moderate, were non-significant. Those correlations were between Item 39 and Item 41 and between Item 40 and Item 41. As a result Item 41 was reworded, as explained earlier, in an attempt to make this item a better fit for the component. In contrast, in the Policies and Procedures component it was found that the correlation between Item 37 and Item 38 was almost perfect ($r=.99$) suggesting that these items are measuring similar aspects. The content of these two items is about having a plan to make policies and procedures, respectively, available to the staff. Therefore, they were combined into one item, considering that one plan including both is sufficient to make evaluation policies and procedures accessible in any organization.

Likewise, inter-item correlations between all the items on the ORECD checklist were also examined. Most of these correlations were positive and more than half were significant showing moderate and high associations. This suggest that, overall, the items on the ORECD checklist are measuring different dimensions of the same construct. Only eight negative inter-item correlations, ranging from $-.01$ to $-.13$, were found. These correlations were non-significant and most of them were among the lowest correlations yielded between all the items. In contrast, the highest correlation ($r=.85$) found within items from different components was the one between Item 12 (Organizational and Leadership Support component) and Item 34 (Communication component) followed by the correlation ($r=.83$) between Item 14 (Organizational and Leadership Support component) and Item 46 (Evaluation Use component). The content of these four items covers aspects of evaluation capacity development that requires the involvement of the organizational leadership, explaining somehow the strength of these correlations.

Limitations of the Study

This study has several limitations to consider because the ORECD checklist is a new instrument that still requires more testing and applications in the field to validate it in a comprehensive way. The experts in Phase 2 volunteered their valuable time meaning that they did not receive any monetary compensation for their work. Due to the limited resources of this study to compensate them for their work and the time constraints to conduct this research, it was not possible to conduct an additional review of the ORECD checklist by experts after the modifications and improvements made based on their initial feedback.

Another limitation was the time in which the interviews with participants in Phase 4 were conducted. Participant interviews were not conducted immediately after receiving their acceptance to participate as is usually done when qualitative data are collected after an event. Therefore, some participants presented difficulties recalling certain facts due to the time passed between the completion of the ORECD checklist and the interview. This may have affected the amount of information or specific details provided by participants.

Although information was gathered about face, content and consequential validity, and the experts provided feedback about how the items in each component fit the component, an additional limitation was that this validation study could not gather information to obtain construct validity of the ORECD checklist. Accordingly, it was listed as one of the aspects that should be examined in future research.

A final limitation was that the scope of this study was confined to the data collected in nonprofit organizations. Nonprofit organizations are just one possible scenario for evaluation capacity development. Including other types of organizations perhaps would have provided a broader perspective of the content of the ORECD checklist. Therefore, the ORECD checklist

must be used carefully with other types of organizations because the nature of nonprofit work (e.g., evaluation requirements from multiple funders), the organizational capacity (i.e., capabilities, knowledge, and resources that nonprofits need in order to be effective), the organizational structure, and context could be different. Evaluation capacity development is context dependent, consequently, any measure developed to be used in one setting possibly needs to be adapted or tailored before using it in a different setting.

Recommendations for Future Research

This study depicted early efforts to evaluate the usefulness of the ORECD checklist using a mixed methods design; therefore, there are several aspects that can be examined in further studies. The recommendations for future research include a field study with nonprofit organizations using the final version of the ORECD checklist produced as a result of the data collected in Phase 4. Also, conducting studies with other types of organizations such as the nonprofit organizations not considered in the current study, for-profit organizations, and government agencies will be useful to inform about the appropriateness and utility of the ORECD checklist for them. These studies could also explore the utility of the ORECD checklist according to the size of the organizations.

Future research will comprise an exploratory factor analysis (EFA) to gather construct validity evidence of the ORECD checklist. This type of factor analysis is indicated because evaluation capacity development is a recent theoretical concept that is still evolving. Therefore, an EFA will allow further exploration of the internal underlying structure of the ORECD checklist to strengthen theory and better understand the ECD multidimensionality. It is important to determine if the current components of the ORECD checklist are the most plausible solution to represent organizational readiness for evaluation capacity development and define if

the items in each component of the ORECD checklist hang together. Moreover, the data collected for the EFA will provide the opportunity to explore the psychometric properties of the ORECD checklist with a larger sample. Furthermore, future research could include a confirmatory factor analysis (CFA) to determine the ability of the factor model obtained in the EFA to fit a specific data set. The CFA will help verify the factor structure suggested by the literature review and used in this study, contributing with valuable information to validate the components of the ORECD checklist and the items in each of the components.

Construct validity evidence of the ORECD checklist can be obtained as well through the development and examination of a nomological network, considering the existing ECD/ECB theory and exploring how the checklist relates to other ECD/ECB and evaluation instruments or variables. This will offer additional understanding on how to evaluate organizational readiness for ECD.

In order to obtain an adequate sample size to conduct the studies recommended in this section, approaching the organizations through the funding sources (e.g., foundations) is recommended. Establishing some type of contact with the organizations prior to sending a research package may help to increase the response rate.

Implications for Practice

The ORECD checklist shows a great potential as a self-assessment instrument to use when stakeholders want to determine the organizational readiness to develop evaluation capacity. It is expected that the ORECD checklist provides organizations insight about the strengths and areas of improvement that need to be prioritized, identifying the extent to which they are present within the organization. The ORECD checklist is suitable as well to identify the progress of the organization toward readiness for ECD by revisiting it as needed. The ORECD checklist can

also be useful to inform organizational consultants who provide evaluation support or coaching to the organizations. Having access to the information on it will facilitate their understanding about where the organization is at any given point and use the results to focus on what needs to be done, to help the organization getting ready to develop evaluation capacity. It is hoped that the ORECD checklist can be a contribution to the body of ECD literature, helping to address organizational readiness in order to support and sustain the development of evaluation capacity and program evaluation efforts within organizations.

Conclusions

This study involved the design of the ORECD checklist and its initial validation. This process provided face, content and consequential validity evidences of the ORECD checklist. It also showed that the ORECD checklist has great potential, establishing its usefulness to determine the readiness of an organization to develop evaluation capacity, as demonstrated by the valuable feedback received from various groups of participants along with its psychometric properties. Overall, the internal consistency results support the reliability of the ORECD checklist. These results along with the correlations suggest that the items are measuring aspects of the same construct. Ongoing efforts should provide additional validity evidences, as the one that can be obtained through a factor analysis, by exploring the underlying structure of the ORECD checklist and its components. Future research will be important as well to explore the utility of the ORECD checklist in organizations other than nonprofits considering that the characteristics and contexts can differ substantially.

REFERENCES

- Adams, J., & Dickinson, P. (2010). Evaluation training to build capability in the community and public health workforce. *American Journal of Evaluation, 31*(3), 421-433.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Long Grove, IL: Waveland Press, Inc.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Arnold, M. E. (2006). Developing evaluation capacity in extension 4-H field faculty: A framework for success. *American Journal of Evaluation, 27*(2), 257-269.
- Atkinson, D. D., Wilson, M., & Avula, D. (2005). A participatory approach to building capacity of treatment programs to engage in evaluation. *Evaluation and Program Planning, 28*, 329-334.
- Baizerman, M., Compton, D. W., & Stockdill, S. H. (2002a). New directions for ECB. *New Directions for Evaluation, 93*, 109-119.
- Baizerman, M., Compton, D. W., & Stockdill, S. H. (2002b). Summary and analysis of the case studies: Themes across the cases. *New Directions for Evaluation, 93*, 101-107.
- Baron, M. E. (2011). Designing internal evaluation for a small organization with limited resources. *New Directions for Evaluation, 132*, 87-99.
- Beere, D. (2005). Evaluation capacity-building: A tale of value-adding. *Evaluation Journal of Australasia, 5*(2), 41-47.
- Behrens, T. R., & Kelly, T. (2008). Paying the piper: Foundation evaluation capacity calls the tune. *New Directions for Evaluation, 119*, 37-50.
- Bickman, L. (1987). The functions of program theory. *New Directions for Program Evaluation, 33*, 5-18.
- BoardSource. (2010). *The handbook of nonprofit governance*. San Francisco, CA: Jossey-Bass.
- Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Sciences, 27*(4), 335-340.

- Boyle, R., & Lemaire, D. (1999). *Building effective evaluation capacity*. New Brunswick, NJ: Transaction Publishers.
- Bozzo, S. L. (2002). Evaluation capacity building in the voluntary/nonprofit sector. *The Canadian Journal of Program Evaluation*, 17(3), 75-92.
- Braverman, M. T., Constantine, N. A., & Slater, J. K. (Eds.). (2004). *Foundations and evaluation: Contexts and practices for effective philanthropy*. Hoboken, NJ: John Wiley & Sons, Inc.
- Brualdi, A. (1999). Traditional and modern concepts of validity. Washington, DC: ERIC Clearinghouse on Assessment and Evaluation (ED435714).
- Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*, 34(3), 299-319.
- Carman, J. G. (2007). Evaluation practice among community-based organizations. *American Journal of Evaluation*, 28(1), 60-75.
- Carman, J. G. (2009). Nonprofits, funders, and evaluation: Accountability in action. *The American Review of Public Administration*, 39(4), 374-390.
- Carman, J. G., & Fredericks, K. A. (2008). Nonprofits and evaluation: Empirical evidence from the field. *New Directions for Evaluation*, 119, 51-71.
- Carman, J. G., & Fredericks, K. A. (2010). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half full? *American Journal of Evaluation*, 31(1), 84-104.
- Carman, J. G., Fredericks, K. A., & Introcaso, D. (2008). Government and accountability: Paving the way for nonprofits and evaluation. *New Directions for Evaluation*, 119, 5-12.
- Clinton, J. (2014). The true impact of evaluation: Motivation for ECB. *American Journal of Evaluation*, 35(1), 120-127.
- Cohen, C. (2006). Evaluation learning circles: A sole proprietor's evaluation capacity-building strategy. *New Directions for Evaluation*, 111, 85-93.
- Compton, D. W. (2009). Managing studies versus managing for evaluation capacity building. *New Directions for Evaluation*, 121, 55-69.
- Compton, D. W., & Baizerman, M. (2007). Defining evaluation capacity building. *American Journal of Evaluation*, 28(1), 118-119.

- Compton, D. W., Baizerman, M., Preskill, H., Rieker, P., & Miner, K. (2001). Developing evaluation capacity while improving evaluation training in public health: the American Cancer Society's collaborative evaluation fellows project. *Evaluation and Program Planning, 24*, 33-40.
- Compton, D. W., Glover-Kudon, R., Smith, I. E., & Avery, M. E. (2002). Ongoing capacity building in the American Cancer Society (ACS) 1995-2001. *New Directions for Evaluation, 93*, 47-61.
- Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australasia, 4*(1-2), 3-11.
- Connolly, P., & Lukas, C. (2002). *Strengthening nonprofit performance: A funders guide to capacity building*. Retrieved from http://www.fieldstonealliance.org/client/article/Article-Components_Org_Capacity.cfm
- Connolly, P., & York, P. (2002). Evaluating capacity-building efforts for nonprofit organizations. *OD Practitioner, 34*(4), 33-39.
- Cousins, J. B., & Bourgeois, I. (2014). Cross-case analysis and implications for research, theory, and practice. *New Directions for Evaluation, 141*, 101-119.
- Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation, 141*, 7-23.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications, Inc.
- Crocker, L., & Algina, J. (2008). *Introduction to classical & modern test theory*. Mason, OH: Cengage Learning.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.
- Dabelstein, N. (2003). Evaluation capacity development: Lessons Learned. *Evaluation, 9*(3), 365-369.
- Díaz-Puente, J. M., Yagüe, J. L., & Alfonso, A. (2008). Building evaluation capacity in Spain. *Evaluation Review, 32*(5), 478-506.
- Duignan, P. (2003). Mainstreaming evaluation or building evaluation capacity? Three key elements. *New Directions for Evaluation, 99*, 7-21.
- Dunkle, T. (2012, January 25). Re: Is your organization accountable and transparent? [Web log message]. Retrieved from <http://bcgcompany.com/blog/nonprofitleader/tag/organization-accountability/>

- Ebrahim, A. (2010). *The many faces of nonprofit accountability*. Working paper 10-069. Retrieved from <http://hbswk.hbs.edu/item/6387.html>
- Evaluation Capacity Development Group. (2009a). What is evaluation capacity development? Retrieved from <http://www.ecdg.net/whatis.htm>
- Evaluation Capacity Development Group. (2009b). Why develop an organization's evaluation capacity? Retrieved from <http://www.ecdg.net/whydevelop.htm>
- Evaluation Capacity Development Group. (2009c). Integrating evaluation into individual jobs. Retrieved from <http://www.ecdg.net/jobdesign.htm>
- Fitzpatrick, J., Sanders, J., & Worthen, B. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Forss, K., Kruse, S. E., Taut, S., & Tenden, E. (2006). Chasing a ghost? An essay on participatory evaluation and capacity development. *Evaluation*, 12(1), 128-144.
- Frechtling, J. A. (2007). *Logic modeling methods in program evaluation*. San Francisco, CA: John Wiley & Sons, Inc.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007) (8th ed.). *Educational research: An Introduction*. Boston, MA: Pearson Education, Inc.
- García-Iriarte, E., Suárez-Balcázar, Y., Taylor-Ritzler, T., & Luna, M. (2011). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation*, 32(2), 168-182
- General Accounting Office. (2003). *Program evaluation: An evaluation culture and collaborative partnerships help build agency capacity*. (GAO-03-454). Retrieved from <http://www.gao.gov/new.items/d03454.pdf>
- Gibbs, D., Napp, D., Jolly, D., Westover, B., & Uhl, G. (2002). Increasing evaluation capacity within community based HIV prevention programs. *Evaluation and Program Planning*, 25, 261-269.
- Gilliam, A., Barrington, T., Davis, D., Lacson, R., Uhl, G., & Phoenix, U. (2003). Building evaluation capacity for HIV prevention programs. *Evaluation and Program Planning*, 26, 133-142.
- George, D., & Mallery, P. (2010). *SPSS for Windows: Step by step* (10th ed.). Boston, MA: Allyn & Bacon.
- Gugiu, P. C., & Rodríguez-Campos, L. (2007). Semi-structured interview protocol for constructing logic models. *Evaluation and Program Planning*, 30, 339-350.

- Hales, B., Terblanche, M., Fowler, R., & Sibbald, W. (2008). Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care*, 20(1), 22-30.
- Hanwright, J., & Makinson, S. (2008). Promoting evaluation culture: The development and implementation of an evaluation strategy in the Queensland Department of Education, Training and the Arts. *Evaluation Journal of Australasia*, 8(1), 20-25.
- Hatcher, L. (1994). *A step-by-step approach to using SAS for factor analysis and structural equation model*. Cary, NC: SAS Institute, Inc.
- Hauge, A. O. (1998). Evaluation capacity development: Lessons learned by the UNDP. In African Development Bank & World Bank, *Evaluation capacity development in Africa: Selected proceedings from a seminar in Abidjan*. Retrieved from [http://lnweb90.worldbank.org/oed/oeddoclib.nsf/a4dd58e444f7c61185256808006a0008/4a9072bd78a8324485256850004fddfc/\\$FILE/Abidjan.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/a4dd58e444f7c61185256808006a0008/4a9072bd78a8324485256850004fddfc/$FILE/Abidjan.pdf)
- Hoole, E., & Patterson, T. E. (2008). Voices from the field: Evaluation as part of a learning culture. *New Directions for Evaluation*, 119, 93-113.
- Huffman, D., Lawrenz, F., Thomas, K., & Clarkson, L. (2006). Collaborative evaluation communities in urban schools: A model of evaluation capacity building for STEM education. *New Directions for Evaluation*, 109, 73-85.
- Huffman, D., Thomas, K., Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation*, 29(3), 358-368.
- Joint Committee on Standards for Educational Evaluation (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Joint Committee on Standards for Educational Evaluation (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Kane, M. (2009). Validating the interpretations and uses of test scores. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 39-64). Charlotte, NC: Information Age Publishing, Inc.
- Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28, 167-172.
- Kapucu, N., Augustin, M. E., & Krause, M. (2007). Capacity building for community-based small nonprofit minority health agencies in central Florida. *The International Journal of Volunteer Administration*, 24(3), 10-17.

- Khan, A. (1998). Evaluation capacity building: An overview of current status, issues and options. *Evaluation, 4*(3), 310-328.
- Khan, A. (2000). Asian perspective in evaluation capacity development: The challenges of the new millennium. In K. Malik & C. Ruth (Eds.), *Evaluation capacity development in Asia* (pp. 8-17). Retrieved from United Nations Development Programme Evaluation Office website: http://www.undp.org/evaluation/documents/ECD/evaluation_cap_dev_china.pdf
- King, J. A. (2002). Building the evaluation capacity of a school district. *New Directions for Evaluation, 93*, 63-64.
- King, J. A. (2005). A proposal to build evaluation capacity at the Bunche-Da Vinci Learning Partnership Academy. *New Directions for Evaluation, 106*, 85-97.
- King, J. A. (2007). Developing evaluation capacity through process use. *Evaluation and Program Planning, 116*, 45-59.
- King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA [Center for Urban and Regional Affairs] Reporter, 35*(3), 10-16.
- 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.
- Kuzmin, A. (2012). Participatory training evaluation method (PATEM) as a collaborative evaluation capacity building strategy. *Evaluation and Program Planning, 35*, 543-546.
- Labin, S. N. (2014). Developing common measures in evaluation capacity building: An iterative science and practice process. *American Journal of Evaluation, 35*(1), 107-115.
- Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation, 33*(3), 307-338.
- LaFond, A., & Brown, L. (2003). *A guide to monitoring and evaluation of capacity building interventions in the health sector in developing countries*. (MEASURE Evaluation Manual Series, No. 7). Retrieved from <http://www.hrhresourcecenter.org/node/1241>
- Lambur, M. T. (2008). Organizational structures that support internal program evaluation. *New Directions for Evaluation, 120*, 41-54.
- Lennie, J. (2005). An evaluation capacity-building process for sustainable community IT initiatives. *Evaluation, 11*(4), 390-414.

- Levin-Rozalis, M., Rosenstein, B., & Cousins, J. B. (2009). A precarious balance: Educational evaluation capacity building in a globalization society. In K. E. Ryan, & J. B. Cousins (Eds.), *The SAGE international handbook of educational evaluation* (pp. 191-212). Thousand Oaks, CA: Sage Publications, Inc.
- Love, A. J. (1983). The organizational context and the development of internal evaluation. *New Directions for Program Evaluation*, 20, 5-22.
- Love, A. J. (1991). *Internal evaluation: Building organizations from within*. Newbury Park, CA: Sage Publications, Inc.
- Love, A. J. (2005). Internal evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Lusthaus, C., Adrien, M. H., & Perstinger, M. (1999). Capacity development: Definitions, issues and implications for planning, monitoring, and evaluation. *Universalia Occasional Paper*, 33, 1-21.
- Mackay, K. (1999). *Evaluation capacity development: A diagnostic guide and action framework* (OED Working Paper Series No. 6.) Retrieved from Impact Alliance website: http://www.impactalliance.org/ev_en.php?ID=1273_201&ID2=DO_TOPIC
- Mackay, K. (2002). The World Bank's ECB experience. *New Directions for Evaluation*, 93, 81-99.
- Majchrzak, A. (1982). Organizational context of program evaluation in community mental health centers. *Evaluation and the Health Professions*, 5(3), 303-333.
- Malik, K. (2000). The new dimensions in evaluation and evaluation capacity development. In K. Malik & C. Ruth (Eds.), *Evaluation capacity development in Asia* (pp. 5-7). Retrieved from United Nations Development Programme Evaluation Office website: http://www.undp.org/evaluation/documents/ECD/evaluation_cap_dev_china.pdf
- Mathison, S. (1991). What do we know about internal evaluations? *Evaluation and Program Planning*, 14, 159-165.
- Mathison, S. (2011). Internal evaluation, historically speaking. *New Directions for Evaluation*, 132, 13-23.
- McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching people to fish? Building the evaluation capability of public sector organizations. *Evaluation*, 9(1), 9-29.
- McGartland Rubio, D., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104.

- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46.
- McMillan, J. H. (2004). *Educational research: Fundamentals for the consumer* (4th ed.). Boston, MA: Pearson Education, Inc.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.
- Milstein, B., Chapel, T.J., Wetterhall, S. F., & Cotton, D. A. (2002). Building capacity for program evaluation at the Centers for Disease Control and Prevention. *New Directions for Evaluation, 93*, 27-46.
- Minnett, A. M. (1999). Internal evaluation in a self-reflective organization: One nonprofit agency's model. *Evaluation and Program Planning, 22*, 353-362.
- Monroe, M. C., Fleming, M. L., Bowman, R. A., Zimmer, J. F., Marcinkowski, T., Washburn, J., & Mitchell, N. J. (2005). Evaluators as educators: Articulating program theory and building evaluation capacity. *New Directions for Evaluation, 108*, 57-71.
- Mott, A. (2003). Hand in hand: Evaluation and organizational development. *The Evaluation Exchange, 9*(3), 8, 12.
- Naccarella, L., Pirkis, J., Kohn, F., Morley, B., Burgess, P., & Blashki, G. (2007). Building evaluation capacity: Definitional and practical implications from an Australian case study. *Evaluation and Program Planning, 30*, 231-236.
- Newcomer, K. E. (2004). How might we strengthen evaluation capacity to manage evaluation contracts? *American Journal of Evaluation, 25*(2), 209-218.
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2004). Meeting the need for practical evaluation approaches: An introduction. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp.xxxiii-xliv). Hoboken, NJ: John Wiley & Sons, Inc.
- Nichols, P. D., & Williams, N. (2009). Consequences of test score use as validity evidence: Roles and responsibilities. *Educational Measurement: Issues and Practice, 28*(1), 3-9.
- Nielsen, S. B., Lemire, S., & Skov, M. (2011). Measuring evaluation capacity: Results and implications of a Danish study. *American Journal of Evaluation, 32*(3), 324-344.
- OECD. (2006). *The challenge of capacity development: Working towards good practice*. Retrieved from http://www.oecd.org/document/30/0,3746,en_2649_33721_3854_9470_1_1_1_1,00.html

- O'Sullivan, R. G., & O'Sullivan, J. M. (1998). Evaluation voices: Promoting evaluation from within programs through collaboration. *Evaluation and Program Planning*, 21, 21-29.
- Owen, J. M. (2003). Evaluation culture: a definition and analysis of its development within organizations. *Evaluation Journal of Australasia*, 3(1), 43-47.
- Owen, J. M. (2005). Organizational culture. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage.
- Owen, J. M. (2007). *Program evaluation: Forms and approaches* (3rd ed.). New York: The Guilford Press.
- Pallant, J. (2010). *SPSS survival manual: A step by step guide to data analysis using the SPSS program* (4th ed.). Chicago, IL: Open University Press.
- Patton, M. Q. (2008). *Utilization-focused evaluation: The new century text* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Personnel. (n.d.). In *Merriam-Webster's Collegiate online dictionary* (11th ed.). Retrieved from <http://www2.merriam-webster.com/cgi-bin/mwdictsn>
- Picciotto, R. (1998). Evaluation capacity development: Issues and challenges. In African Development Bank & World Bank, *Evaluation capacity development in Africa: Selected proceedings from a seminar in Abidjan*. Retrieved from [http://lnweb90.worldbank.org/oed/oeddoclib.nsf/a4dd58e444f7c61185256808006a0008/4a9072_bd78a8324485256850004fddfc/\\$FILE/Abidjan.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/a4dd58e444f7c61185256808006a0008/4a9072_bd78a8324485256850004fddfc/$FILE/Abidjan.pdf)
- Preskill, H. (2008). Evaluation's second act: A spotlight on learning. *American Journal of Evaluation*, 29(2), 127-138.
- Preskill, H. (2014). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, 35(1), 116-119.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.
- Preskill, H., & Russ-Eft, D. (2005). *Building evaluation capacity; 72 activities for teaching and training*. Thousand Oaks, CA: Sage Publications, Inc.
- Reckase, M. D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, 17(2), 13-16.
- Rodríguez-Campos, L., & Rincones-Gómez, R. (2013). *Collaborative evaluations: Step by step* (2nd ed.). Stanford, CA: Stanford University Press.

- Rotondo, E. (2012). Lessons learned from evaluation capacity building. *New Directions for Evaluation*, 134, 93-101.
- Russ-Eft, D., & Preskill, H. (2009). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change* (2nd ed.). New York, NY: Basic Books.
- Russon, K., & Russon, C. (2007). *Evaluation Capacity Development Group toolkit* (2nd ed.). Retrieved from <http://www.ecdg.net/ordering-toolkit-english.htm>
- Satterlund, T. D., Treiber, J., Kipke, R., Kwon, N., & Cassady, D. (2013). Accommodating diverse clients' need in evaluation capacity building: A case study of the tobacco control evaluation center. *Evaluation and Program Planning*, 36, 49-55.
- Schaumburg-Müller, H. (1996). *Evaluation capacity building: Donor support and experiences*. Report for the Development Assistance Committee Expert Group on Aid Evaluation, OECD. Retrieved from <http://www.oecd.org/dataoecd/20/52/16546669.pdf>
- Schwandt, T. A. (2008). Educating for intelligent belief in evaluation. *American Journal of Evaluation*, 29(2), 139-150.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage Publications, Inc.
- Scriven, M. (2005). Checklists. In S. Mathison (Ed.), *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage
- Scriven, M. (2007). *The logic and methodology of checklists*. Retrieved from http://www.wmich.edu/evalctr/archive_checklists/papers/logic&methodology_dec07.pdf
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428.
- Simister, N., & Smith, R. (2010). *Monitoring and evaluating capacity building: Is it really that difficult?* (Praxis Paper 23). Retrieved from International NGO Training and Research Centre website: <http://www.intrac.org/resources.php?action=resource&id=677>
- Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 19-37). Charlotte, NC: Information Age Publishing, Inc.
- Sobeck, J., & Agius, E. (2007). Organizational capacity building: Addressing a research and practice gap. *Evaluation and Program Planning*, 30, 237-246.
- Sonnichsen, R. C. (1999). Building evaluation capacity within organizations. In R. Boyle, & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from practice* (pp. 53-73). New Brunswick, New Jersey: Transaction Publishers.

- Sridharan, S., & De Silva, S. (2010). On ambition, learning, and co-evolution: Building evaluation as a field. *American Journal of Evaluation*, 31(2), 246-251.
- Staff. (n.d.). In *Merriam-Webster's Collegiate online dictionary* (11th ed.). Retrieved from <http://www2.merriam-webster.com/cgi-bin/mwdictsn>
- Stevenson, J. F., Florin, P., Mills, D. S., & Andrade, M. (2002). Building evaluation capacity in human service organizations: a case study. *Evaluation and Program Planning*, 25, 233-243.
- 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.
- Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: A practical guide to their development and use* (4th ed). New York: Oxford University Press Inc.
- Stufflebeam, D. L. (2000). *Guidelines for developing evaluation checklists: The checklists development checklist*. Retrieved from http://www.wmich.edu/evalctr/archive_checklists/guidelines_cdc.pdf
- Stufflebeam, D. L. (2001). Evaluation checklist: Practical tools for guiding and judging evaluations. *American Journal of Evaluation*, 22(1), 71-79.
- Stufflebeam, D. L. (2002). *Institutionalizing evaluation checklist*. Retrieved from http://www.wmich.edu/evalctr/archive_checklists/institutionalizingeval.pdf
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, & applications*. San Francisco, CA: John Wiley & Sons, Inc.
- Suárez-Balcázar, Y., & Harper, G. W. (2003). Building capacity for participatory evaluation within community initiatives. *Journal of Prevention & Intervention in the Community Monographic Separates*, 26(2), 21-36.
- Suárez-Balcázar, Y., & Taylor-Ritzler, T. (2014). Moving from science to practice in evaluation capacity building. *American Journal of Evaluation*, 35(1), 95-99.
- Taut, S. (2007a). Defining evaluation capacity building: Utility considerations. *American Journal of Evaluation*, 28(10), 120.
- Taut, S. (2007b). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation*, 28(1), 45-59.
- Taveras, S., Duncan, T., Gentry, D., Gilliam, A., Kimbrough, I., & Minaya, J. (2007). The evolution of the CDC HIV prevention capacity-building assistance initiative. *Journal of Public Health Management and Practice*, S8-S15.

- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. *New Directions for Evaluation*, 120, 55-69.
- Taylor-Ritzler, T., Suárez-Balcázar, Y., García-Iriarte, E., Henry, D. B., & Balcázar, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation*, 34(2), 190-206.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp.3-50). Thousand Oaks, CA: Sage Publications, Inc.
- Trevisan, M. S. (2002). Evaluation capacity in K-12 school counseling programs. *American Journal of Evaluation*, 23(3), 291-305.
- Tseng, S. H. (2011). An assessment scale of ECB: Perspectives from the faculty and staff of technical colleges in Taiwan. *Evaluation and Program Planning*, 34, 154-159.
- United Way of America. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: United Way of America.
- UNDP. (2006). *Capacity development*. Practice Note. Retrieved from http://europeandcis.undp.org/uploads/public/File/Capacity_Development_Regional_Training/UNDP_Capacity_Development_Practice_Note_JULY_FINAL.pdf
- Valéry, R., & Shakir, S. (2005). Evaluation capacity building and humanitarian organization. *Journal of Multidisciplinary Evaluation*, 3, 78-111.
- Volkov, B. B. (2011). Internal evaluation a quarter-century later: A conversation with Arnold J. Love. *New Directions for Evaluation*, 132, 5-12.
- Volkov, B. B., & Baron, M. E. (2011). Issues in internal evaluation: Implications for practice, training, and research. *New Directions for Evaluation*, 132, 101-111.
- Volkov, B. B., & King, J. A. (2007). *A checklist for building organizational evaluation capacity*. Retrieved from http://www.wmich.edu/evalctr/archive_checklists/ecb.pdf
- Wandersman, A. (2014). Moving forward with the science and practice of evaluation capacity building (ECB): The why, how, what, and outcomes of ECB. *American Journal of Evaluation*, 35(1), 87-89.
- Weinberg, G. M. (2001). *An introduction to general systems thinking*. Broadway, NY: Dorset House Publishing.
- Wisely, D. S. (2002). Parting thoughts on foundation evaluation. *American Journal of Evaluation*, 23(2), 159-164.

Yaghmale, F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3(1), 25-27.

Zhaoying, C. (2003) *Donor support for evaluation capacity development: From the perspective of recipient-country*. Report for DAC/WP-EV Workshop. Retrieved from http://www.minefi.gouv.fr/fonds_documentaire/TRESOR/cicid/atelier/contrib/43pdf

APPENDICES

Appendix A: Initial Version of the ORECD Checklist

Organizational Readiness for Evaluation Capacity Development Checklist				
Connie Walker-Egea 2012				
<p>This checklist is based on a review of relevant literature regarding evaluation capacity building and evaluation capacity development. The purpose of the checklist is to guide stakeholders in determining the readiness of their organization to develop or strengthen its evaluation capacity. The checklist serves the organizations to: (a) identify their current situation to support the development of evaluation capacity; (b) guide them in recognizing which areas are in need of improvement; and (c) determine their progress toward readiness for evaluation capacity development, by revisiting the checklist when necessary.</p>				
For each of the following statements, circle the response that best describe the current situation of your organization.	To a Great Extent	To Some Extent	To a Small Extent	Not at All
A. Organizational Environment				
1. Organizational environment allows the establishment/strengthening of an internal evaluation system.	4	3	2	1
2. Organization is aware of the benefits of internal evaluation.	4	3	2	1
3. Role of evaluation is understood by the whole organization (not only the leadership).	4	3	2	1
4. Evaluation is valued by the whole organization.	4	3	2	1
5. Evaluation is considered a means to provide important information.	4	3	2	1
6. Personnel are committed to developing the evaluation capacity of the organization.	4	3	2	1
7. Personnel understand how evaluation can contribute to organizational learning.	4	3	2	1
8. Personnel are committed to learning from evaluation.	4	3	2	1
9. Evaluators are perceived as facilitators.	4	3	2	1
10. Evaluators are perceived as a useful resource.	4	3	2	1
11. Personnel understand the importance of incorporating evaluation into everyday work practices.	4	3	2	1
12. Personnel know how their work relates to evaluation.	4	3	2	1
13. External stakeholders (e.g., funders) support the development of evaluation capacity.	4	3	2	1
B. Organizational Leadership Support				
14. Leadership supports the development of internal evaluation capacity.	4	3	2	1
15. Leadership has a long-term commitment to support ongoing evaluation capacity development.	4	3	2	1

Appendix A (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
16. Leadership encourages involvement of personnel in the development of evaluation capacity.	4	3	2	1
17. Leadership supports practices that integrate evaluation into organizational life.	4	3	2	1
18. Leadership supports the integration of evaluation at the decision-making level.	4	3	2	1
19. Leadership is committed to the development of an infrastructure to support the development of evaluation capacity.	4	3	2	1
20. Leadership is committed to the development of an evaluative learning culture.	4	3	2	1
C. Knowledge/Skills Development				
21. Organization is aware of the current evaluation skills/knowledge of the personnel.	4	3	2	1
22. There is a mechanism to identify evaluation capacity needs of the personnel.	4	3	2	1
23. Organization is committed in providing professional development to increase evaluation skills/knowledge of the personnel.	4	3	2	1
24. Training opportunities at different organizational levels are available to support the continuity of evaluation capacity efforts.	4	3	2	1
25. Evaluation expertise is available inside the organization.	4	3	2	1
26. Personnel with experience conducting evaluations are available.	4	3	2	1
D. Resources				
27. Organization has sufficient <u>human</u> resources to develop evaluation capacity.	4	3	2	1
28. Organization has sufficient <u>financial</u> resources to develop evaluation capacity.	4	3	2	1
29. Evaluation budget is a priority for the organization.	4	3	2	1
30. Evaluation budget reflects the evaluation needs of the organization.	4	3	2	1
31. Evaluation budget is consistent with organization's long term objectives.	4	3	2	1
32. Appropriate tools/technologies are available (e.g., computer hardware and software, equipment, materials).	4	3	2	1
33. Personnel have sufficient time during the workday to collaborate on evaluation activities.	4	3	2	1
E. Program Theory				
34. Personnel are aware of how organizational program(s) work.	4	3	2	1
35. Personnel know how to construct logic models.	4	3	2	1
36. Logic model(s) is used as an evaluation planning tool (define and clarify what and when to evaluate).	4	3	2	1

Appendix A (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
F. Demand for Evaluation				
37. There is demand for information (e.g., on what works, accountability requirements, strategies to increase evaluation use) from external stakeholders.	4	3	2	1
38. There is demand to develop evaluation capacity.	4	3	2	1
39. There is commitment to develop evaluation capacity.	4	3	2	1
40. There is a plan to develop evaluation capacity.	4	3	2	1
G. Communication				
41. An effective organizational communication system is in place.	4	3	2	1
42. Organization has communication structures to facilitate the flow of information across the organization (i.e., informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	4	3	2	1
43. Organization has communication procedures to manage information (e.g., collection and dissemination).	4	3	2	1
H. Policies and Procedures				
44. Organization has evaluation/evaluation capacity development <u>policies</u> in place (e.g., rules to guide evaluation decisions and actions, evaluation training for personnel, ethical considerations, disclosure of information).	4	3	2	1
45. Organization has evaluation/evaluation capacity development <u>procedures</u> in place (e.g., handbooks and manuals, sources of and guidelines to obtain technical assistance in evaluation, consultation processes).	4	3	2	1
46. Personnel are aware of the evaluation/evaluation capacity development <u>policies</u> .	4	3	2	1
47. Personnel are aware of the evaluation/evaluation capacity development <u>procedures</u> .	4	3	2	1
I. External Support				
48. Organization collaborates with external evaluation experts to support the development of evaluation capacity.	4	3	2	1
49. Organization shares evaluation knowledge/skills with other organizations to support the development of evaluation capacity.	4	3	2	1
50. Organization employs mechanisms for advancing evaluation capacity (e.g., building networks and relationships, identifying outstanding practices from successful organizations to use them as standards for comparison).	4	3	2	1

Appendix A (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
J. Incentives				
51. Incentives are available to encourage participation in the development of evaluation capacity (e.g., time and flexibility to incorporate the evaluation process into daily work practices).	4	3	2	1
52. Incentives are available to encourage involvement of personnel in evaluation.	4	3	2	1
53. Personnel are acknowledged for their contributions to evaluation.	4	3	2	1
54. Personnel understand how evaluation can contribute to their individual learning (e.g., development of valuable, lifelong skills).	4	3	2	1
K. Feedback Mechanism				
55. Organization has a feedback mechanism.	4	3	2	1
56. Personnel are receptive to feedback from others.	4	3	2	1
57. Feedback mechanism is part of the decision-making process.	4	3	2	1
L. Evaluation Use				
58. Organizational leadership supports the use of evaluative information.	4	3	2	1
59. Dissemination of evaluation results is promoted.	4	3	2	1
60. Personnel are able to use evaluation results (e.g., for planning, decision-making).	4	3	2	1
61. Evaluation results are used for continuous improvement of the organization.	4	3	2	1

Appendix B: Expert Review Form

Part 1. Please rate the relevance of each item to determine organizational readiness for evaluation capacity development and the clarity of each item. The items are presented by components known to contribute to developing evaluation capacity within organizations. At the end of each component please provide feedback about whether the items fit the component.

Item	Relevance to Content Area				Clear/Free of Ambiguity			
	1	2	3	4	1	2	3	4
A. Organizational Environment: Understanding of the internal and external organizational contexts, organizational culture, and everyday ways of working to determine whether building evaluation capacity is feasible.								
1. Organizational environment allows the establishment/strengthening of an internal evaluation system.								
2. Organization is aware of the benefits of internal evaluation.								
3. Role of evaluation is understood by the whole organization (not only the leadership).								
4. Evaluation is valued by the whole organization.								
5. Evaluation is considered a means to provide important information.								
6. Personnel are committed to developing the evaluation capacity of the organization.								
7. Personnel understand how evaluation can contribute to organizational learning.								
8. Personnel are committed to learning from evaluation.								
9. Evaluators are perceived as facilitators.								
10. Evaluators are perceived as a useful resource.								
11. Personnel understand the importance of incorporating evaluation into everyday work practices.								
12. Personnel know how their work relates to evaluation.								
13. External stakeholders (e.g., funders) support the development of evaluation capacity.								
Please comment about the extent to which items 1-13 fit the organizational environment component.								
B. Organizational Leadership Support: Engagement and support to sustain evaluation capacity development and integrate evaluation into organizational life.								
14. Leadership supports the development of internal evaluation capacity.								
15. Leadership has a long-term commitment to support ongoing evaluation capacity development.								
16. Leadership encourages involvement of personnel in the development of evaluation capacity.								

Appendix B (Continued)

Item	Relevance to Content Area				Clear/Free of Ambiguity			
	1	2	3	4	1	2	3	4
17. Leadership supports practices that integrate evaluation into organizational life.								
18. Leadership supports the integration of evaluation at the decision-making level.								
19. Leadership is committed to the development of an infrastructure to support the development of evaluation capacity.								
20. Leadership is committed to the development of an evaluative learning culture.								
Please comment about the extent to which items 14-20 fit the organizational leadership support component.								
C. Knowledge/Skills Development: Development of some level of evaluation expertise within the organization at the individual and organizational level and at different organizational levels.								
21. Organization is aware of the current evaluation skills/knowledge of the personnel.								
22. There is a mechanism to identify evaluation capacity needs of the personnel.								
23. Organization is committed in providing professional development to increase evaluation skills/knowledge of the personnel.								
24. Training opportunities at different organizational levels are available to support the continuity of evaluation capacity efforts.								
25. Evaluation expertise is available inside the organization.								
26. Personnel with experience conducting evaluations are available.								
Please comment about the extent to which items 21-26 fit the knowledge/skills development component.								
D. Resources: Availability of organizational resources such as financial, personnel, time, tools, and technology to sustain evaluation capacity development and evaluation practices.								
27. Organization has sufficient <u>human</u> resources to develop evaluation capacity.								
28. Organization has sufficient <u>financial</u> resources to develop evaluation capacity.								
29. Evaluation budget is a priority for the organization.								
30. Evaluation budget reflects the evaluation needs of the organization.								
31. Evaluation budget is consistent with organization's long term objectives.								
32. Appropriate tools/technologies are available (e.g., computer hardware and software, equipment, materials).								

Appendix B (Continued)

Item	Relevance to Content Area				Clear/Free of Ambiguity			
	1	2	3	4	1	2	3	4
33. Personnel have sufficient time during the workday to collaborate on evaluation activities.								
Please comment about the extent to which items 27-33 fit the resources component.								
E. Program Theory: Understanding of program theory and logic model as a way to develop evaluation capacity within the organization.								
34. Personnel are aware of how organizational program(s) work.								
35. Personnel know how to construct logic models.								
36. Logic model(s) is used as an evaluation planning tool (define and clarify what and when to evaluate).								
Please comment about the extent to which items 34-36 fit the program theory component.								
F. Demand for Evaluation: Demand for information from external stakeholders and demand to develop evaluation capacity.								
37. There is demand for information (e.g., on what works, accountability requirements, strategies to increase evaluation use) from external stakeholders.								
38. There is demand to develop evaluation capacity.								
39. There is commitment to develop evaluation capacity.								
40. There is a plan to develop evaluation capacity.								
Please comment about the extent to which items 37-40 fit the demand for evaluation component.								
G. Communication: Existence of an organizational communication system.								
41. An effective organizational communication system is in place.								
42. Organization has communication structures to facilitate the flow of information across the organization (i.e., informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).								
43. Organization has communication procedures to manage information (e.g., collection and dissemination).								
Please comment about the extent to which items 41-43 fit the communication component.								

Appendix B (Continued)

Item	Relevance to Content Area				Clear/Free of Ambiguity			
	1	2	3	4	1	2	3	4
H. Policies and Procedures: Development of policies and procedures for evaluation and evaluation capacity development.								
44. Organization has evaluation/evaluation capacity development <u>policies</u> in place (e.g., rules to guide evaluation decisions and actions, evaluation training for personnel, ethical considerations, disclosure of information).								
45. Organization has evaluation/evaluation capacity development <u>procedures</u> in place (e.g., handbooks and manuals, sources of and guidelines to obtain technical assistance in evaluation, consultation processes).								
46. Personnel are aware of the evaluation/evaluation capacity development <u>policies</u> .								
47. Personnel are aware of the evaluation/evaluation capacity development <u>procedures</u> .								
Please comment about the extent to which items 44-47 fit the policies and procedures component.								
I. External Support: Support and collaboration from other organizations, agencies, and external evaluators.								
48. Organization collaborates with external evaluation experts to support the development of evaluation capacity.								
49. Organization shares evaluation knowledge/skills with other organizations to support the development of evaluation capacity.								
50. Organization employs mechanisms for advancing evaluation capacity (e.g., building networks and relationships, identifying outstanding practices from successful organizations to use them as standards for comparison).								
Please comment about the extent to which items 48-50 fit the external support component.								
J. Incentives: Availability of incentives to encourage evaluation capacity development.								
51. Incentives are available to encourage participation in the development of evaluation capacity (e.g., time and flexibility to incorporate the evaluation process into daily work practices).								
52. Incentives are available to encourage involvement of personnel in evaluation.								
53. Personnel are acknowledged for their contributions to evaluation.								

Appendix B (Continued)

Item	Relevance to Content Area				Clear/Free of Ambiguity			
	1	2	3	4	1	2	3	4
54. Personnel understand how evaluation can contribute to their individual learning (e.g., development of valuable, lifelong skills).								
Please comment about the extent to which items 51-54 fit the incentives component.								
K. Feedback Mechanism: Existence of a feedback mechanism to facilitate the learning process in the organization.								
55. Organization has a feedback mechanism.								
56. Personnel are receptive to feedback from others.								
57. Feedback mechanism is part of the decision-making process.								
Please comment about the extent to which items 55-57 fit the feedback mechanism component.								
L. Evaluation Use: Use of evaluation findings for different organizational purposes.								
58. Organizational leadership supports the use of evaluative information.								
59. Dissemination of evaluation results is promoted.								
60. Personnel are able to use evaluation results (e.g., for planning, decision-making).								
61. Evaluation results are used for continuous improvement of the organization.								
Please comment about the extent to which items 58-61 fit the evaluation use component.								

Part 2. Please provide feedback about the items of the checklist and overall checklist. Please write your comments in this form.

- Does the checklist appear to be a good measure of the elements necessary for evaluation capacity development? Please explain.
- What changes to the checklist, if any, do you consider to be necessary?
- Are the items clear? If no, what needs to be changed?
- Is the response scale appropriate? If no, what needs to be changed?
- Would you revise any item(s)? If yes, please explain.
- Are there any items that should be added to the checklist? If yes, please explain.
- Are there any items that should be deleted from the checklist? If yes, please explain.

Appendix C: Professional Background Questionnaire

Please answer the following questions in the space provided. All responses will be kept confidential and reported only in aggregate.

1. What is the highest academic degree you have received?
 Doctoral degree
 Master's degree
 Bachelor's degree
 Other (please indicate): _____
2. How many years of experience do you have in the evaluation field? Please provide a whole number.

_____ years
3. How many years have you worked in evaluation capacity building/development? Please provide a whole number.

_____ years
4. Are you currently active in evaluation capacity building/development?
 Yes
 No
5. Do you have experience developing checklists?
 Yes
 No

Appendix D: Pretesting Version of the ORECD Checklist

Organizational Readiness for Evaluation Capacity Development Checklist				
Connie Walker-Egea 2012				
<p>This checklist is based on a review of relevant literature regarding evaluation capacity building and evaluation capacity development. The goal of evaluation capacity development (ECD) is to put in place and sustain the components that support program evaluation efforts within the organization. The purpose of the checklist is to guide stakeholders in determining the readiness of their organization to develop or strengthen its evaluation capacity. The checklist serves the organizations to: (a) identify their current situation to support the development of evaluation capacity; (b) guide them in recognizing which areas are in need of improvement; and (c) determine their progress toward readiness for evaluation capacity development, by revisiting the checklist when necessary.</p>				
Directions: For each of the following statements, circle the response that best describe the current situation of your organization, indicating the extent to which it is present within the organization. After completing the checklist, review the statements marked as “to a small extent” and “not at all”. These are the areas of improvement that need to be prioritized to support the development evaluation capacity.	To a Great Extent	To Some Extent	To a Small Extent	Not at All
A. Organizational Environment				
1. Internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).	4	3	2	1
2. Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).	4	3	2	1
3. There is a general understanding of how evaluation can provide important information to the organization.	4	3	2	1
4. There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.	4	3	2	1
5. There is a general commitment to learning from evaluation throughout the organization	4	3	2	1
6. Organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.	4	3	2	1
7. Staff is aware of how their work relates to evaluation.	4	3	2	1
8. There is organizational stability (for example, organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and staff turnover is low).	4	3	2	1
B. Organizational Leadership Support				
9. Leadership supports the development of internal evaluation capacity.	4	3	2	1

Appendix D (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
10. Leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).	4	3	2	1
11. Leadership is committed to encouraging evaluation capacity development activities.	4	3	2	1
12. Leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.	4	3	2	1
13. Leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).	4	3	2	1
C. Knowledge/Skills Development				
14. Leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.	4	3	2	1
15. There is a plan to identify the strengths and weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).	4	3	2	1
16. There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work).	4	3	2	1
17. There is staff with evaluation expertise/experience conducting evaluations inside the organization to support the ECD process.	4	3	2	1
D. Resources				
18. There are strategies in place to be able to access sufficient <u>human</u> resources for the ECD process.	4	3	2	1
19. Organization has committed <u>financial</u> resources to develop evaluation capacity.	4	3	2	1
20. There is a plan for securing additional fiscal resources to develop evaluation capacity, as needed.	4	3	2	1
21. Evaluation budget is a priority for the organization.	4	3	2	1
22. There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.	4	3	2	1
23. There is a plan to provide staff sufficient time during the workday to work on evaluation activities.	4	3	2	1
E. Program Theory				
24. Organizational program(s) goals/objectives are well defined.	4	3	2	1
25. Staff has a common understanding about <u>what</u> organizational program(s) do.	4	3	2	1

Appendix D (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
26. Staff has a common understanding of <u>how</u> organizational program(s) work (for example, there is a logic model or any other graphical representation of the program(s) and the expected outcomes).	4	3	2	1
F. Demand for Evaluation				
27. There is demand for evaluative information from <u>external</u> stakeholders (for example, on what works, accountability requirements, and strategies to increase evaluation use).	4	3	2	1
28. There is demand for evaluative information from <u>internal</u> sources.	4	3	2	1
29. There is a written plan about how to develop evaluation capacity.	4	3	2	1
G. Communication				
30. There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).	4	3	2	1
31. There are communication <u>structures</u> to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	4	3	2	1
32. There are communication <u>procedures</u> to manage/share information (such as, collection, dissemination, and disclosure of information).	4	3	2	1
33. There is a feedback mechanism (action or means used to modify the ECD process as a result of information received) to facilitate learning within the evaluation capacity development process.	4	3	2	1
34. Leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.	4	3	2	1
H. Policies and Procedures				
35. There is a plan in place to develop ECD <u>policies</u> (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).	4	3	2	1
36. There is a plan in place to develop ECD <u>procedures</u> (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).	4	3	2	1
37. There is a plan to make ECD <u>policies</u> accessible to all staff.	4	3	2	1
38. There is a plan to make ECD <u>procedures</u> accessible to all staff.	4	3	2	1

Appendix D (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
I. External Support				
39. External stakeholders (for example, funders) support the development of evaluation capacity.	4	3	2	1
40. Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.	4	3	2	1
41. There is a plan to establish mechanisms for advancing the development of evaluation capacity (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).	4	3	2	1
J. Incentives				
42. Incentives are available to encourage staff participation in the development of evaluation capacity (for example, time and flexibility people need to incorporate evaluation into the everyday work of the organization).	4	3	2	1
43. There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.	4	3	2	1
44. Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).	4	3	2	1
K. Evaluation Use				
45. Leadership promotes the dissemination of evaluation results (to inform staff, as appropriate, about them).	4	3	2	1
46. Leadership supports the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).	4	3	2	1
47. Leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).	4	3	2	1

Appendix E: Questions for Pretesting

1. Are the items clearly worded?
 - a. If the items are not clearly worded, what needs to be changed to improve them?

2. Do the items flow in a logical order?
 - a. If the flow of the items does not have a logical order, what needs to be changed to improve it?

3. Is the physical format of the checklist appropriate for its purpose?
 - a. If the physical format is not appropriate, what needs to be changed to improve it?

Appendix F: Field Study Version of the ORECD Checklist

Organizational Readiness for Evaluation Capacity Development Checklist				
Connie Walker-Egea 2012				
<p>This checklist is based on a review of relevant literature regarding evaluation capacity building and evaluation capacity development. The goal of evaluation capacity development (ECD) is to put in place and sustain the components that support program evaluation efforts within the organization. The purpose of this checklist is to guide stakeholders in determining the readiness of their organization to develop or strengthen its evaluation capacity. The checklist serves organizations by: (a) identifying the current situation to support the development of evaluation capacity, (b) guiding in recognizing which areas may be in need of improvement, and (c) determining progress toward readiness for evaluation capacity development by revisiting the checklist when necessary.</p>				
<p>Directions: For each of the following statements, circle the response that best describes the current situation of your organization, indicating the extent to which it is present within the organization. Notice that some statements provide examples to facilitate answering them but these examples are by no means exhaustive. After completing the checklist, review the statements marked as “to a small extent” and “not at all.” These are the areas of improvement that need to be prioritized in order to support the development of evaluation capacity.</p>	To a Great Extent	To Some Extent	To a Small Extent	Not at All
A. Organizational Environment				
1. The internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).	4	3	2	1
2. Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).	4	3	2	1
3. There is a general understanding of how evaluation can provide important information to the organization.	4	3	2	1
4. There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.	4	3	2	1
5. There is a general commitment to learning from evaluation (process and/or results) throughout the organization	4	3	2	1
6. The organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.	4	3	2	1
7. Staff is aware of how their work relates to evaluation.	4	3	2	1
8. There is organizational stability (for example, the organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, and the staff turnover is low).	4	3	2	1

Appendix F (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
B. Organizational Leadership Support				
9. The organizational leadership supports the development of internal evaluation capacity.	4	3	2	1
10. The organizational leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).	4	3	2	1
11. The organizational leadership is committed to encouraging evaluation capacity development activities.	4	3	2	1
12. The organizational leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.	4	3	2	1
13. The organizational leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).	4	3	2	1
14. The organizational leadership is committed to investing in training/professional development to increase evaluation skills/knowledge of the staff involved in ECD.	4	3	2	1
C. Knowledge/Skills Development				
15. There is a plan to identify the strengths/weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).	4	3	2	1
16. There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work vs. traditional lectures).	4	3	2	1
17. There is staff with evaluation expertise/experience conducting evaluations inside the organization in order to support the ECD process.	4	3	2	1
D. Resources				
18. There are strategies in place to be able to access sufficient human resources for the ECD process.	4	3	2	1
19. Organization has committed financial resources to develop evaluation capacity.	4	3	2	1
20. There is a plan for securing additional fiscal resources to develop evaluation capacity.	4	3	2	1
21. An evaluation budget is a priority for the organization.	4	3	2	1
22. There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.	4	3	2	1
23. There is a plan to provide staff sufficient time during the workday to work on evaluation activities.	4	3	2	1

Appendix F (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
E. Program Theory				
24. The goals/objectives of the organizational program(s) are well defined.	4	3	2	1
25. Staff has a common understanding about what organizational program(s) do.	4	3	2	1
26. Staff has a common understanding of how organizational program(s) work (for example, there is a logic model or other graphical representation of the program(s) and the expected outcomes).	4	3	2	1
F. Demand for Evaluation				
27. There is demand for evaluative information from external stakeholders, such as funders (for example, on what works, accountability requirements, and strategies to increase evaluation use).	4	3	2	1
28. There is demand for evaluative information from internal sources.	4	3	2	1
29. There is a written plan about how to develop evaluation capacity.	4	3	2	1
G. Communication				
30. There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).	4	3	2	1
31. There are communication structures to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	4	3	2	1
32. There are communication procedures to manage/share information (such as, collection, dissemination, and disclosure of information).	4	3	2	1
33. There is a feedback mechanism (action or means used to modify the ECD process as a result of information received) to facilitate learning within the evaluation capacity development process.	4	3	2	1
34. The organizational leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.	4	3	2	1
H. Policies and Procedures				
35. There is a plan in place to develop ECD policies (including for example, rules to guide ECD decisions, actions, and activities; evaluation training for staff; ethical considerations; and disclosure of information).	4	3	2	1
36. There is a plan in place to develop ECD procedures (these are step by step instructions to put policies in action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in evaluation and consultation processes).	4	3	2	1
37. There is a plan to make ECD policies accessible to all staff.	4	3	2	1
38. There is a plan to make ECD procedures accessible to all staff.	4	3	2	1

Appendix F (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
I. External Support				
39. External stakeholders (for example, funders) support the development of evaluation capacity.	4	3	2	1
40. Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.	4	3	2	1
41. There is a plan to establish mechanisms for advancing the development of evaluation capacity (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).	4	3	2	1
J. Incentives				
42. Incentives are available to encourage staff participation in the development of evaluation capacity (for example, allotted time and flexibility for people to incorporate evaluation into the everyday work of the organization).	4	3	2	1
43. There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.	4	3	2	1
44. Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).	4	3	2	1
K. Evaluation Use				
45. The organizational leadership promotes the dissemination of evaluation results (to inform staff about them as appropriate).	4	3	2	1
46. The organizational leadership promotes the use of evaluative information (for example, to internally monitor program activities and understand what is working or not).	4	3	2	1
47. The organizational leadership/staff is able to use evaluation results (for example, for planning, decision-making, when deciding how to implement, deliver, and improve programs, and when identifying lessons about what has been effective).	4	3	2	1

Appendix H: Nonprofit Organization Demographics

Code _____

Please answer the following questions about the organization. Some questions may require more than one answer:

1. Year the organization was founded:

2. Annual operating budget:
 Up to \$25, 000
 \$25,001 - \$50,000
 \$50, 001 - \$100,000
 \$100,001 - \$250,000
 \$250,001 - \$500,000
 \$500,001 - \$750,000
 \$750,001 - \$1million
 More than \$1 million
3. Number of employees:
_____ Full-time employees
_____ Part-time employees
4. The organization offers services in the following area(s) (Mark **ALL** that apply):
 Educational
 Health Care
 Human Services
 Youth Development
 Other: _____
5. Funding sources of the organization (Mark **ALL** that apply):
 Local government
 State government
 Federal government
 Foundation
 Other: _____
6. Who is primarily responsible for the evaluation? (Mark **ALL** that apply):
 Internal evaluator
 External evaluator
 Evaluation unit within the organization
 Organizational staff **with** evaluation training
 Organizational staff **without** evaluation training
 Other: _____
 Organization **does not** conduct evaluation
7. Describe the evaluation activities of the organization:
 We do not conduct evaluations
 We do very little evaluation
 We do some evaluation
 We make a great effort to evaluate our program(s)
 We go above and beyond to evaluate our program(s)
8. When you completed this checklist you were:
 Thinking about the entire organization
 Thinking about a department or unit in the organization
9. How many people were in charge of completing the checklist?

Please indicate your availability to participate in a follow-up interview.

Yes, I would like to participate

No, I would not like to participate

Appendix I: Field Study Interview Protocol

Time of Interview: _____

Date: _____

Organization: # _____

Job Title of Interviewee: _____

Questions:

1. Can you describe the process you followed while completing the checklist?
 - a. Did you experience any difficulties when completing the checklist?
 - b. How time consuming was to complete the checklist?
 - c. How helpful was the physical format of the checklist?
2. Do you foresee any benefits for your organization as a result of using this checklist? Why or why not?
 - a. Which aspects of the checklist are most beneficial?
 - b. To what extent do you think these benefits are likely to be realized?
3. Do you foresee any issues or problems for your organization as a result of using this checklist? Why or why not?
 - a. Which aspects of the checklist are most problematic?
 - b. To what extent do you think these issues or problems are likely to be realized?
4. Is there something else that you would like to add about your experience using the checklist?
5. Are there any further recommendations for improvement of the checklist?

Appendix J: Final Version of the ORECD Checklist

Organizational Readiness for Evaluation Capacity Development Checklist				
Connie Walker-Egea 2014				
<p>This checklist is based on a review of relevant literature regarding evaluation capacity building and evaluation capacity development. The goal of evaluation capacity development (ECD) is to put in place and sustain the components that support program evaluation efforts within the organization. The purpose of this checklist is to guide stakeholders in determining the readiness of their organization to develop or strengthen its evaluation capacity. The checklist serves organizations by: (a) identifying the current situation to support the development of evaluation capacity, (b) guiding in recognizing which areas may be in need of improvement, and (c) determining progress toward readiness for evaluation capacity development by revisiting the checklist when necessary.</p>				
<p>Directions: For each of the following statements, circle the response that best describes the current situation of your organization, indicating the extent to which it is present within the organization. Notice that some statements provide examples to facilitate answering them but these examples are by no means exhaustive. After completing the checklist, review the statements marked as “to a small extent” and “not at all.” These are the areas of improvement that need to be prioritized in order to support the development of evaluation capacity.</p>	To a Great Extent	To Some Extent	To a Small Extent	Not at All
A. Organizational Environment				
1. The internal organizational environment allows the establishment/strengthening of an evaluation system to support formal evaluations (for example, staff is ready and willing to receive new ideas, has positive attitudes toward evaluation, and there are rewards for innovation and creativity).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Staff is aware of the benefits of conducting internal evaluation (for example, staff understands the role of evaluation and values its contributions).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. There is a general understanding of how evaluation can provide important information to the organization.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. There is a general understanding of how evaluation can contribute to organizational learning throughout the organization.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. There is a general commitment to learning from evaluation (process and/or results) throughout the organization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. The organization has identified evaluation champions who are committed to evaluation (with time and ability), to help lead/sustain the ECD process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Staff is aware of how their work relates to evaluation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. There is organizational stability (some aspects to consider are: the organization has clearly defined and commonly understood vision and mission, has a clear direction about where is going in the near and distant future, has decision-making procedures, has control of the finances, and the staff turnover is low).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix J (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
B. Organizational Leadership Support				
9. The organizational leadership supports the development of internal evaluation capacity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. The organizational leadership is committed to supporting ongoing evaluation capacity development (for example, devoting resources and infrastructure/foundation necessary for this process).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. The organizational leadership is committed to encouraging evaluation capacity development activities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. The organizational leadership is committed to supporting practices that integrate evaluation into the ongoing work of the organization.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. The organizational leadership is committed to the development of an evaluative learning culture (for example, using evaluative information to support and challenge the work of the organization, making time to learn, and learning from mistakes and experiences).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. The organizational leadership is committed to investing in training/professional development (including for example, types of evaluation approaches and methods of data collection) to increase the evaluation skills/knowledge of the staff involved in ECD.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. The organizational leadership has a written plan about how to develop evaluation capacity in the organization.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C. Knowledge/Skills Development				
16. There is a plan to identify the strengths/weaknesses of the staff regarding evaluation (for example, a needs assessment of their current evaluation skills/knowledge and areas for development).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. There is a plan to develop staff skills/knowledge using strategies that engage people in collaborative learning in addition to the traditional formal presentations (examples of collaborative learning are: mutual learning, knowledge transfer, learning by doing, mentorship, and paired work vs. traditional lectures).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. There is staff inside the organization with evaluation expertise/experience conducting evaluations in order to support the ECD process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
D. Resources				
19. There are strategies in place to be able to access sufficient human resources for the ECD process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. Organization can commit financial resources to develop evaluation capacity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. There is a plan for securing fiscal resources to develop evaluation capacity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. An evaluation budget is a priority for the organization.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. There is a plan for accessing appropriate tools/technologies (such as, computer hardware and software, equipment, and materials) to support the ECD process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. There is a plan to provide staff sufficient time during the workday to work on ECD/evaluation activities.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix J (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
E. Program Theory				
25. The goals/objectives of the organizational program(s) are well defined.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26. Staff has a common understanding about what organizational program(s) do.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. Staff has a common understanding of how organizational program(s) work (for example, there is a logic model or other graphical representation of the program(s) and the expected outcomes).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F. Demand for Evaluation				
28. There is demand for evaluative information from external stakeholders, such as funders (for example, on what works, accountability requirements, and strategies to increase evaluation use).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. There is demand for evaluative information from internal sources (for example, to increase funding sources, for program improvement).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. There is demand to develop evaluation capacity (either internal/external source).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
G. Communication				
31. There is an effective communication system (for example, it allows sending and receiving information for decision making and problem-solving).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
32. There are communication structures to facilitate the flow of information across the organization (that is, informal, formal, vertical-between different levels of authority, and horizontal-between people on the same level of authority).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
33. There are communication procedures to manage/share information (such as, collection, dissemination, and disclosure of information).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
34. There is a feedback mechanism (action or means used to modify a process as a result of information received) to facilitate learning within the evaluation capacity development process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
35. The organizational leadership is willing to receive feedback from others (such as staff and external evaluator) in order to increase the impact of the evaluation capacity development process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
H. Policies and Procedures				
36. There is a plan in place to develop evaluation policies (including for example, rules to guide evaluation decisions, actions and activities; evaluation training for staff; ethical considerations; and disclosure of information).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
37. There is a plan in place to develop evaluation procedures (these are step by step instructions to put policies into action, including for example, handbooks and manuals, and guidelines to obtain technical assistance in any ECD and evaluation consultation processes).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
38. There is a plan to make evaluation policies/procedures accessible to all staff.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix J (Continued)

	To a Great Extent	To Some Extent	To a Small Extent	Not at All
I. External Support				
39. External stakeholders (for example, funders) support the development of evaluation capacity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
40. Staff is willing to collaborate with external evaluation experts to support the development of evaluation capacity.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
41. There is a written plan to establish mechanisms for advancing the development of evaluation capacity by joining efforts with external sources (for example, collaborating, building networks and sharing knowledge and experiences with external partners; participating in communities of practice, and identifying outstanding practices from successful organizations to use them as standards for comparison).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
J. Incentives				
42. Incentives are available to encourage staff participation in the development of evaluation capacity (for example, allotted time and flexibility for people to incorporate evaluation into the everyday work of the organization).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
43. There is a plan to acknowledge staff contributions (individual and group contributions) to the development of evaluation capacity of the organization.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
44. Staff is aware of how their participation in the development of evaluation capacity can contribute to their individual learning (for example, development of valuable, lifelong skills).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
K. Evaluation Use				
45. The organizational leadership has identified means to promote the dissemination of evaluation results as appropriate (for example, meetings, reports, newsletters).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
46. The organizational leadership has identified means to promote the use of evaluative information (for example, giving access to the information, translating the results into appropriate action, and exploring the positive and negative implications of using the results).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
47. The organizational leadership/staff is willing to make use of evaluation results (for example, for planning, decision-making, deciding how to implement, deliver, and improve programs, and identifying what has been effective).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix K: Approval Letter from the Institutional Review Board



DIVISION OF RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
13901 Bruce B. Downs Blvd. MDC035 • Tampa, FL 33612-4199
(813) 974-5638 • FAX (813) 974-5616

May 10, 2012

Connie Walker-Egea, M.A.
Edu Measurement & Research
4204 E. Fowler Ave., EDU 105
Tampa, FL 33620

RE: **Expedited Approval** for Initial Review
IRB#: Pro00008122
Title: Design and Validation of an Evaluation Checklist for Organizational Readiness for
Evaluation Capacity Development

Dear Ms. Walker-Egea:

On 5/9/2012 the Institutional Review Board (IRB) reviewed and **APPROVED** the above referenced protocol. Please note that your approval for this study will expire on 5/9/2013.

Approved Items:

Protocol Document:

[Design and Validation of an Evaluation Checklist for Organizational Readiness for Evaluation Capacity Development](#)

Consent Document:

Your study qualifies for a waiver of the requirements for the documentation of informed consent as outlined in the federal regulations at 45CFR46.117 (c) which states that an IRB may waive the requirement for the investigator to obtain a signed consent form for some or all subjects if it finds either: (1) that the only record linking the subject and the research would be the consent document and the principal risk would be potential harm resulting from a breach of confidentiality. Each subject will be asked whether the subject wants documentation linking the subject with the research, and the subject's wishes will govern; or (2) that the research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context.

It was the determination of the IRB that your study qualified for expedited review which includes activities that (1) present no more than minimal risk to human subjects, and (2) involve only procedures listed in one or more of the categories outlined below. The IRB may review research through the expedited review procedure authorized by 45CFR46.110 and 21 CFR 56.110. The research proposed in this study is categorized under the following expedited review

categories:

(6) Collection of data from voice, video, digital, or image recordings made for research purposes.

(7) Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

As the principal investigator of this study, it is your responsibility to conduct this study in accordance with IRB policies and procedures and as approved by the IRB. Any changes to the approved research must be submitted to the IRB for review and approval by an amendment.

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

Sincerely,



John A. Schinka, Ph.D., Chairperson
USF Institutional Review Board