Development and Validation of an Evaluation Capacity Instrument for the Early Childhood Field

by

Btissam El Hassar

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Measurement, Evaluation and Cognition

Department of Educational Psychology
University of Alberta

Abstract

This study contributes to the conceptualization and measurement of evaluation capacity (EC) by developing and validating an EC instrument and an empirically based conceptual EC framework that is relevant to the context of the early childhood field in Alberta, Canada. As evaluation has become a crucial strategy for addressing learning and accountability demands faced by organizations, a common challenge for organizations is how to access adequate capacity for conducting evaluations and meaningfully using the results. Previous research has defined and conceptualized the construct of EC, but less attention has been given to its measurement, especially measurement that is grounded in context. This gap warranted a new examination to assess the factors that contribute to building effective EC within specific contexts, the relationships among the factors, and the expected outcomes of EC efforts. The need for better measurement of EC is evidenced by numerous calls from scholars and practitioners as recently as 2017.

A cross-sectional survey design was used to answer the study's research questions about the underlying factor structure of the EC construct, the empirical relationships among these factors, and the influence of participants' organizational roles on the EC relationships. The study's research approach conforms to the contemporary view of validity, in which multiple types of evidence are produced during the validation process to substantiate the validity argument. The research began by drawing on the literature and consulting experts who are knowledgeable about the study's context to identify and operationalize relevant EC components and develop an EC instrument that is empirically sound and contextually relevant. Data were then collected from participants in the early childhood field using the instrument.

Exploratory and confirmatory factor analyses were used to evaluate measurement models that were based on individual and organizational EC scales. These results suggest that both individual and organizational EC constructs are multidimensional and that they are adequately measured by the developed EC instrument. Structural equation modelling was then used to evaluate direct and indirect relationships among the individual and organizational factors that were hypothesized within the EC framework. The direct and indirect relationships were found to be statistically significant. Specifically, the latent variables Organizational Culture, Organizational Leadership, Organizational Commitment to Evaluation, and Individual Evaluation Skills, and Individual Attitude Toward Evaluation positively predicted Individual Motivation to Conduct and Use Evaluations. Furthermore, these relationships explained 49% of the variance in Individual Motivation. Subsequently, multi-group path analysis was used to assess the influence of respondents' roles within their organizations on these relationships. The multi-group path analysis results revealed that respondents' professional roles did not statistically influence the direct and indirect EC relationships.

This study offers important implications for theory, methodology, and practice. EC theory is advanced through the contribution of an empirical EC framework specifying the magnitude and directionality of the influences among individual and organizational EC factors. In so doing, the framework highlights the interrelatedness of the factors and thus the importance of adopting a systems perspective to inform EC conceptualization, measurement and initiatives. A context-relevant methodology contributes to generating valid conclusions about the extent to which EC exists within individuals and organizations. In so doing, the methodology provides an illustrative example for scholars to apply the study procedures in contexts beyond the early childhood field. Finally, the instrument enhances the capacity for practitioners to access data

that may inform EC building initiatives. In this way, the instrument can be used by individuals and organizations for evidence-based decision-making that targets specific EC needs.

Preface

This thesis is an original work by Btissam El Hassar. The research project of which this thesis is part of received research ethics approval from the University of Alberta Research Ethics Board, Project Name "EVALUATION CAPACITY NETWORK," No. Pro00048028, 2015-01-23. The dissertation research has been funded by the generosity of Social Sciences and Humanities Research Council of Canada (SSHRC) and the Women and Children's Health Research Institute (WCHRI).

Acknowledgements

I would like to express my deepest gratitude and appreciation to the many people whose support, encouragement and friendship were invaluable before and during my doctoral journey. They enriched my experience and made it possible for me to succeed. They helped me to persevere when things got tough, and taught me to appreciate the process.

I am very grateful first to Dr. Cheryl N. Poth, my doctoral supervisor, for her incredible support, feedback, and advice, which helped me think critically about my work. I appreciate her providing me with opportunities take part in research and evaluation projects and connecting me to international scholars during conferences. Her mentorship went beyond my dissertation document. When I felt barely above water, Cheryl was there with words of wisdom and kindness that helped me find perspective.

A special thanks to Dr. Rebecca Gokiert and Dr. Okan Bulut, who served on my doctoral dissertation committee. Working with Dr. Gokiert closely as a research assistant made it possible for me to learn from and contribute to her work with communities and organizations in the early childhood development field. She encouraged me to think of usefulness and application of my own research. Likewise, Dr. Bulut's guidance as I made decisions about approaches to quantitative analysis was invaluable. In addition, I would like to extend thanks to Dr. Stanley Varnhagen and Dr. Jill Chouinard, members of my examining committee, who offered many insightful comments and questions.

I would like also to thank the expert reviewers, the project management team and the core research team for their time, questions, and feedback. Their contributions have been invaluable. A sincere thank you as well to Dr. Mark Gierl, the director of CRAME, who set the stage for my doctoral journey. He was the first person to officially welcome me to our program and to explain

in an empathetic, a calm and warm way the challenges and opportunities ahead. I learned a great deal from him as a teacher and a scholar, and I always enjoyed our conversations. I also appreciate the support of Dr. Ying Cui, who recommended I take Dr. Poth's program evaluation course and the mentorship of Dr. Jacquie Pei while working with her on evaluation projects as a member of ACCERT.

I feel fortunate to have smart, dedicated, and fun colleagues in CRAME and ACCERT. I appreciate our collaborations and friendships. A special thanks to Qi Guo for his support with my dissertation research and for our many conversations about politics, AI, and culture, which I enjoyed very much.

My doctoral journey was enriched and positively impacted by the support I received and the relationships I built with the non-academic staff of our department, and in the Student Success Center. While writing a dissertation can be a lonely journey, being part of the Thesis Retreat and the Write-In gave me a community with a sense of a shared purpose.

It is also important to acknowledge that my doctoral journey was possible because of the influence that many people had on me even before I started it.

I am deeply grateful to Mary Carol Dragoo and the late Alan Dragoo, who invited me to undertake undergraduate studies in the United States and supported my study. They connected me with the Seekers community, who showed me what openness, acceptance, and kindness mean; their actions and commitments are not limited by borders, ethnicity, or religion. I am grateful for all their love, support, inspiration, and mentorship. I also want to thank all my host families and my mentors for all their love and wisdom.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Adjusted BIC   Bayesian information criterion adjusted for model overfitting

AIC   Akaike information criterion

BIC   Bayesian information criterion

CFA   Confirmatory factor analysis

CFI   Comparative fit index

df   Degrees of freedom

EC   Evaluation capacity

EFA   Exploratory factor analysis

MLR   Full-information maximum likelihood with robust standard error

RMSEA   Root mean square error of approximation

SEM   Structural equation modeling

SRMR   Standardized root mean square residual

TLI   Tucker-Lewis index

Chapter 1: Introduction

This dissertation study bridges the fields of evaluation, measurement, and research methods to generate timely and contextually relevant theoretical, methodological, and practical contributions. Drawing on the existing literature across these fields highlighted the need to advance an empirical conceptualization and measurement of evaluation capacity (EC), and an ongoing project provided the opportunity for this study within the context of the early childhood field in Alberta, Canada. To that end, this study advances an empirical conceptualization and measurement of EC by developing a contextually relevant instrument to assess individual and organizational EC that is consistent with the contemporary views and practices of measurement and validation.

In this chapter, I begin by framing the study within the field of evaluation and explaining the need for a contextually relevant conceptualization and measurement of EC. I then explain the study context related to the EC gap within the early childhood field and the relationship of this study to an existing project titled "Evaluation Capacity Network: Aligning Evaluative Thinking and Practice among Early Childhood Stakeholders." Following that, the purpose of this dissertation is articulated, followed by a discussion of my philosophical perspectives as a researcher to describe how my views may have influenced my choices and decisions during the process. The chapter closes by outlining the overall structure of the dissertation.

Framing the Study

The term *evaluation* evokes a multiplicity of meanings, as a result, there is no one definition that all scholars agree on (Alkin & King, 2017; King & Sevahn, 2013). Initially, evaluation scholars focused on an aspect of evaluation that is derived from the term, which was rooted in determining "value" or "worth" (Alkin, 2018). This focus raised questions about the

objectives of evaluation, its purpose(s), the perspective it should adopt, and the methods that can be selected (Alkin, 2018). The concept of evaluation has expanded from this initial narrow conception to one that reflects diverse philosophical views and methodological approaches (Alkin & Christie, 2004; Christie & Alkin, 2008). This expansion has made it more difficult to settle on a common definition. The lack of consensus may be explained by the diversity of evaluators' practice areas (e.g., health sciences, economics, or education) and their theoretical views on the purpose of evaluation (e.g., use or methods; Christie & Alkin, 2008). In this study, I adopt a definition that reflects my view that the evaluation approach chosen should focus on maximizing use of the findings by stakeholders. Patton's (2011a) definition of program evaluation is consistent with this view because he defines evaluation as the "systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding" (Patton, 2011a, p. 35). This definition emphasizes the importance of attending to the needs of multiple stakeholders. Evaluators often have to balance the needs of stakeholders representing a variety of organizations, ranging from community-based organizations to government agencies (Meadows, 1998; Patton, 2011a; Taut, 2007). By focusing on meeting the needs of multiple stakeholders, program evaluation has the potential to provide information that meets diverse organizational needs, such as monitoring and reporting requirements of funders, organizational learning, and public policy (Carman, 2007).

When combined, human interactions inherent within organizations, the social issues organizational members aim to address, and the economic and political milieu within which organizations exist create a dynamic environment (Chouinard, 2013; Patton, 2011a). These

dynamic environments highlight the need for organizations to assess their goals, processes, and impacts for the purposes of learning and/or accountability. To address these learning challenges and accountability demands, organizations turn to evaluation because of the flexibility inherent in the multiple approaches and methodologies it offers (Shaw, Greene, & Mark, 2006). The need for learning and accountability evaluation can originate either within or outside of the organization. To address organizational learning needs, evaluations can be used to inform strategic decision-making, improve operations and practices, and encourage organizational members to think evaluatively—all with the aim of embedding learning and adapting practices within the organization (Patton, 2011a). To address organizational accountability needs, evaluations are used to satisfy funding requirements, inform policymaking, and decide on the allocation of funds (Patton, 2011a; Wilcox & King, 2013). Frequently, the reasons for conducting evaluations are not dichotomous but in fact integrate elements of both learning and accountability (Chelimsky, 2006).

Evaluators require specific skills, resources, and other supports to plan and conduct an evaluation that meets the intended purposes (CES, 2009). These requirements generate at least three major challenges for organizations. First, they may not have adequate capacity to conduct evaluations and use the results meaningfully (e.g., Preskill & Boyle, 2008a; Suarez-Balcazar et al., 2013). Few organizations have access to the skills, resources, and/or supports required for undertaking meaningful evaluations, let alone funding or time to develop the skills needed to meet evolving accountability and learning objectives (Cheng & King, 2016; Janzen et al., 2017). Second, organizations and their funders may find that their evaluation needs are misaligned (Carman & Fredericks, 2010; Liket, Rey-Garcia, & Maas, 2014). Funders, for example, request evaluations that can be used as management tools for decision-making purposes, which are not

necessarily useful to community organizations (Chouinard, 2013). With their limited EC, organizations who prioritize evaluation for accountability will do so at the cost of failing to conduct evaluations that address their learning needs (Biott & Cook, 2000; Chouinard, 2013).

Finally, organizations may lack access to the expertise needed to guide them in undertaking evaluations. Conducting evaluations requires organizations to either hire external evaluation consultants, who are costly, or to build capacity within the organization itself. Both options—external and internal—are difficult because of limited funding and other capacity constraints (Bakken, Núñez, & Couture, 2014). While use of an external evaluator may seem to be the easier choice, organizations still need to develop internal capacity in order to hire someone who is qualified, oversee the evaluation process, and use the evaluation results (Arnold, 2006; Cousins, Goh, Elliott, & Bourgeois, 2014; Preskill & Boyle, 2008). For many organizations, the sustainable option is to build internal EC that meets organizational needs, which is understandably not straightforward (e.g., Cousins et al., 2014; Preskill & Boyle, 2008). Thus, both the external and the internal options require organizations to have sufficient in-house EC. The efforts and processes that are involved in building EC bring about the very capabilities that are needed in evaluation (Bourgeois & Cousins, 2013; Cousins, 2004). Undertaking EC building can thus be considered an investment in organizational time and resources.

As is the case with any activity that requires financial and human investment, it is necessary to determine the contributing factors that make EC building efforts successful and to ascertain the potential outcomes of those efforts (e.g., Mark, Greene, & Shaw, 2006). Building EC effectively and efficiently within organizations requires an empirical understanding of EC conceptualization and measurement. Moreover, any attempt to build EC must also consider the influence of context on this effort (e.g., Mark et al., 2006; Suarez-Balcazar et al., 2010), because

context may play an important role in shaping individual and organizational EC needs and, therefore, should shape the initiatives to address those EC needs.

The Need to Conceptualize and Measure Evaluation Capacity

The goal of building EC is to provide organizations and their members with the necessary skills and resources to conduct and use evaluations, but, as simple as this goal may sound, EC is challenging to define, conceptualize, and measure (Cousins et al., 2014; Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012; Preskill & Boyle, 2008). Even though much effort has been devoted to defining the construct of EC (and EC building) and describing the components that comprise EC, there remains some ambiguity (e.g., Labin et al., 2012; Preskill & Boyle, 2008). For example, EC and EC building are often used interchangeably (Cheng & King, 2016). Although the two terms are related, EC refers to an existing level of capacity that may be an outcome of efforts made to improve it, while the EC building refers to a process that includes a set of strategies targeting aspects of EC (Cheng & King, 2016). Additionally, only a few studies have empirically considered the context in which conceptualizing, measuring, and building EC take place. Ignoring context limits our understanding of both the process and the outcomes of the EC endeavour (Preskill & Boyle, 2008; Suarez-Balcazar et al., 2013). Having a limited understanding of EC theory and measurement make organization's efforts to improve the effectiveness and efficiency of their own EC building strategies challenging. It is difficult to address these challenges because there is a gap in the literature regarding which specific components contribute to building EC, how to measure those components, and which relationships among the components are important.

The process of conceptualizing and measuring EC is complex because of the potential interactions and the dynamic environments stemming from the human elements, the organizational culture, the disciplinary fields, and the programmatic contexts (e.g., Fierro, 2012; Mark et al., 2006; Preskill & Boyle, 2008a). In addition, the social, economic, and political environments that encompass an organization and influence its functioning, add to the complexity involved in understanding EC (Patton, 2011). Defining, conceptualizing, and measuring EC, therefore, requires a methodological approach that considers the EC components, their conceptualization and measurement that is context-specific (Stockdill, Baizerman, & Compton, 2002; Suarez-Balcazar et al., 2010; Wandersman, 2014).

It is possible to use a systems approach as a lens for understanding EC. A systems approach entails moving away from a narrow view in order to understand a complex issue by capturing its essential parts and the relationships among them (Patton, 2011a; Williams & Imam, 2007). With this lens, evaluators' skills are understood as only one aspect of the overall system. In the absence of a systems perspective, organizations may overlook the importance of other parts of the system, such as the organization's motivation for evaluation, the organizational learning culture, the leadership style, and the organizational context. Understanding EC from a systems perspective requires a contextually relevant empirical understanding of the EC construct that balances comprehensiveness and specificity.

Focusing on the theoretical conceptualization and measurement of EC in a specific context can help organizations improve the effectiveness and efficiency of their own EC building efforts. Being able to conceptualize and measure the components of EC can help organizations prioritize their EC efforts based on evidence regarding how best to build EC (Preskill, 2013). Furthermore, measuring EC can make it possible for organizations to evaluate their efforts and

understand their dynamic needs. Much of the limited research in the EC area has focused more on defining and conceptualizing EC or EC building in general, and less on its measurement and contextual relevance (Labin, 2014; Preskill, 2013; Suarez-Balcazar et al., 2013). Therefore, it is necessary to create an instrument for measuring EC and establishing the validity and reliability of interpretations within the intended EC context.

Evaluation Capacity within the Study Context

This dissertation study is part of a four-year project called *Evaluation Capacity Network: Aligning Evaluative Thinking and Practice among Early Childhood Stakeholders* (henceforth, *the overall project*), which was funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). The project was influenced by dynamics stemming from the number of stakeholders involved, the social and political environment characterizing the areas of early childhood and evaluation, and the fragmentation of the early childhood field (Alberta Government, 2013). This section first describes the need for EC in the early childhood field and then briefly describes the overall project to provide the intended context for the EC framework and instrument.

Evaluation capacity in the early childhood field. The process of developing into a healthy adult begins during one's early years (birth to six years) and requires the involvement of multiple disciplines, sectors, and a variety of programs (Shonkoff & Phillips, 2000). The importance of early childhood experiences in laying the foundation for later health and well-being is well established in the literature (Masten & Gewirtz, 2006; Shonkoff & Phillips, 2000; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2010; Trawick-Smith, 2013). In Alberta, approximately one in four children experience difficulties and fall below the Canadian norm across five areas of development: physical health and well-being, social competence,

emotional maturity, language and thinking skills, and communication skills and general knowledge (Early Childhood Mapping Project, 2014), and the challenges faced by Alberta are not unique. The reality faced by children in Alberta has increased the importance of and need for evidence-informed early childhood development programs (Shonkoff, 2017; Shonkoff & Phillips, 2000). Gathering useful evidence to inform these programs requires organizations to have the capacity to conduct evaluation and use its findings.

Addressing organizations' needs for EC in this context is challenging because the early childhood field involves multiple sectors and is influenced by the economic, social, and political environment (Ferns & Friendly, 2012; Langford & Richardson, 2018). In the multisectoral early childhood field in Canada, organizations must often deal with multiple funding agencies that often have different accountability requirements (Gokiert et al., 2017). These challenges are magnified by limited funding and available human resources. Organizations, therefore, struggle when investing their limited resources in evaluation, which often does not go beyond meeting the accountability requirement (Shonkoff, 2017). Addressing organizations' needs for developing their EC requires a deeper understanding of the challenges they face, along with their capacity gaps and opportunities. To that end, the Evaluation Capacity Network (Network) was created to facilitate and contribute sustainable solutions.

The relationship of this study to an existing project. The purpose of the overall project was to develop an intersectoral and multidisciplinary partnership—a network—with the goal of advancing evaluation practices in the field of early childhood in Alberta, Canada. This was accomplished through building EC, including enhancing evaluative and reflective thinking and practice, creating space for dialogue and a shared understanding of the evaluation-related challenges faced across the diverse early childhood field, and proposing potential sustainable

solutions. The Network is a partnership of diverse members from government agencies, community organizations, funding organizations, and academic institutions at the local, provincial, and national levels. These organizations address different aspects of early childhood development (e.g., early learning and care, health, and education), which makes their needs diverse but also interconnected (Muijs, Aubrey, Harris, & Briggs, 2004).

The Network's governance includes three teams: a project management team and two advisory bodies (the core research team and the steering committee). The advisory bodies advise the project management team on management and research decisions for meeting the partnership's goals. Collectively, the advisory bodies provide diverse expertise, including expertise in measurement and evaluation, early childhood development, research methods, and survey design. They also represent the diversity of the Network's early childhood stakeholders, including academic institutions, community organizations, governmental bodies, and funding agencies. To achieve its goal, the Network engaged in three primary activities: (1) assessing the EC needs of early childhood organizations, (2) developing opportunities for EC, and (3) supporting ongoing dialogue among early childhood stakeholders about improving evaluation practices.

The present dissertation study was embedded in the Network development and activities. Working closely with members of the project management team, the core research team, and the steering committee in the development process, I integrated multiple data sources to develop a context-relevant EC instrument. The members of these teams were a valuable resource, given their expertise in key areas of this study, including the early childhood context. I also gathered primary data from early childhood organizations for use in the EC instrument's validation process. It was very beneficial to have my study embedded in the overall project for several

reasons, including having regular access early on to experts regarding the conceptualization and measurement of EC within the study's context as well as having access to potential respondents.

Purpose of this Dissertation

This study aims to advance an empirical conceptualization and measurement of EC by developing an EC instrument that is relevant to the early childhood field, and that follows a rigorous methodology to enhance the validity of the inferences made from the data. The EC instrument is to be used to assess the existing individual and organizational evaluation capacity at the aggregate level. Using a self-administered instrument facilitated the collection of data from a wide range of respondents within the early childhood field in a timely and cost-effective manner (Blair et al., 2013). The study purpose was achieved by building on existing theoretical and empirical knowledge; consulting experts with diverse knowledge of the study context and the areas of evaluation, measurement, and methodology; and analyzing the psychometric properties of the EC instrument using factor analysis and structural equation modelling. Ultimately, the goal of contributing an empirically based and contextually relevant framework and instrument for assessing EC is to enable organizations to make informed decisions about which aspects of their EC require improvement and conduct periodic re-assessments of those components.

Personal Positionality

The philosophical perspectives that underpin my approach as a researcher have been shaped by my personal and educational journey. They were informed by the axiological, ontological, and epistemological assumptions reflected in my areas of research and in the methodological approach I selected for this dissertation study. My views are consistent with a pragmatic perspective, that is, a paradigm that values plurality and places more importance on

answering the questions that have been asked than on the methods used to answer them (e.g., Crotty, 1998; Johnson, Onwuegbuzie, & Turner, 2007). In this section, I briefly describe my journey to elucidate the influences on my philosophical perspectives, using an organizing structure based on the philosophical assumptions.

Fundamental to the research process are the researcher's ontological views regarding the nature of reality. A pragmatic paradigm holds that the nature of reality can be both singular—in the sense that a phenomenon can be explained using a single theory— and pluralistic, requiring different perspectives to understand a phenomenon (Creswell & Plano-Clark, 2011). As in the case of my axiological views, my values, personal, and educational experiences have influenced my ontological views. I have studied diverse topics (e.g., economics, policy, evaluation) in different countries, a cultural experience that has allowed me to meet colleagues and friends from different regions of the world who have their own cultural, political, and economic realities. Being introduced to such a diversity of perspectives has challenged me to expand my views. This diversity has taught me to be sensitive to the different assumptions and worldviews held by others—an insight that applies to good research as well as to social relationships. This learning has been further enhanced by my experience in evaluation, which has helped me to appreciate the value of using multiple methods (e.g., expert reviews, statistical methods) during the development of an EC instrument to gather diverse evidence to answer evaluation questions pertaining to stakeholders with varying perspectives.

In order to abide by one's axiological and ontological views, one must adopt a research approach that reflects these. In accordance with my pragmatic values and view of reality, my epistemological assumptions prompted me to balance the need to answer the research questions, to employ a diverse methodology, and to ensure feasibility. In particular, I used multiple data

sources to gain knowledge about how to conceptualize and measure EC components related to the context of the overall project (i.e., the early childhood field). My ultimate goal in developing and using the EC instrument was to provide early childhood programs with an instrument that can produce useful information about their EC to help them strategically target EC gaps by conducting meaningful evaluations for diverse purposes (e.g., organizational learning, accountability).

Outline of the Dissertation

This dissertation consists of five chapters: an introduction to the research (Chapter 1) that lays out the motivations for the study and the study context; a literature review (Chapter 2) that integrates the literature pertaining to the theoretical and methodological rationales for this study; a methodology section (Chapter 3) that describes the procedures involved in the development and validation of the EC framework and instrument; a section that presents the study's findings (Chapter 4), reports the results addressing the study's research questions; and finally, a discussion (Chapter 5) that delineates the study's interpretations, its theoretical, methodological, and practical contributions, its limitations and suggestions for future research, and closes with an overall conclusions.

Chapter 2: Literature Review

Given that many organizations lack the necessary capacity for conducting evaluations and using the results, EC has emerged as an important construct that researchers have explored both theoretically and empirically (e.g., Cheng & King, 2016; Labin et al., 2012; Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry, & Balcazar, 2013). The interest in building and measuring EC has spawned the creation of instruments developed to capture its components (Goh, Quon, & Cousins, 2007; Nielsen, Lemire, & Skov, 2011; Preskill & Torres, 2001; Taylor-Ritzler et al., 2013). One of the pressing challenges for researchers is accessing adequate data for assessing the validity of these instruments (Labin, 2014). Contemporary notions of validity call for embedding efforts to generate validity evidence throughout the process of instrument development (Zumbo & Chan, 2014).

To that end, this study adopts an approach that reflects the unified view of validity (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) within the process of instrument development. This chapter is organized in four sections. The first section, on the potential of the unified view of validity during instrument development, explains the role of validation processes and distinguishes the unified view of validity from traditional approaches. The second section argues for the importance of embedding the validation in the process of instrument development by highlighting the current lack of a theoretical foundation in EC measurement. The third section establishes the need to focus on developing a contextually relevant instrument that integrates the unified view of validity. The fourth section summarizes the existing EC conceptualizations and instruments that contributed to the development of this

study's instrument and framework (described in Chapter 3). The chapter ends by describing the need for the study and delineating the research questions guiding the study.

The Role of Validation in Instrument Development

Measurement instruments, also known in the literature as surveys or tests, are tools that researchers use to collect data (AERA et al., 2014). Instruments are designed to address research questions about relationships among indicators, the predictability of outcomes, and comparisons between groups (Wiersma & Jurs, 2009). Validity is defined as "the degree to which evidence and theory support the interpretation of test [instrument] scores for proposed uses of tests [instruments]" (AERA et al., 2014, p. 11). There is a consensus that the notion of validity has to do with the construct being measured (e.g., EC in this study) and whether the inferences that are drawn from the data are supported by evidence (Cronbach & Thorndike, 1971; Messick, 1989; Zumbo & Chan, 2014).

As numerous professional and scholarly organizations have attested, pursuing efforts to enhance the validity of inferences is "the most fundamental consideration in developing and evaluating tests" (AERA et al., 2014, p. 9). Valid interpretations of the data generated from an instrument can, for example, be used in evaluations to effectively inform program and policy decisions that potentially have far-reaching impacts (Cizek, 2012). The use of an instrument that lacks supporting validity evidence, by contrast, may generate conclusions that have unintended negative ramifications for those using the data to make decisions (Kane, 2013a). Making decisions based on weak validity evidence may therefore not only be unethical, but may also negatively impact many people, organizations, and policies (Abma, 2006; Davies, Newcomer, & Soydan, 2006; Moss, 1998). Establishing the validity of instruments is especially relevant for

evaluators who tend to work in multidisciplinary sectors, in which the necessary instruments are not always readily available (Weitzman & Silver, 2013).

A comprehensive validity framework provides a foundation that can guide the process of creating an instrument and generating multiple lines of validity evidence to substantiate potential inferences (AERA et al., 2014). There is strong agreement among validity theorists that the process of validation needs to be grounded in theory and the context in which the instrument will be used (Chan, 2014; Moss, 1998; Zumbo, 2009). Such a process begins by defining the construct to be measured, a step that is undertaken even before creating the items and that continues throughout the development and use of the instrument (Zumbo, 2009). The process should also include specification of the validation approach to be followed and a description of the procedures that generate the validity evidence (AERA et al., 2014; Zumbo & Chan, 2014).

Despite consensus among validity scholars about the validation process and recommended practices, these recommendations are often not followed in practice (Collie & Zumbo, 2014; Hubley, Zhu, Sasaki, & Gadermann, 2014; Kane, 2001). A collection of systematic reviews focusing on validation practice revealed that this oversight even occurs in prominent measurement journals such as *Educational and Psychological Measurement* (Shear & Zumbo, 2014). It is noteworthy that current instruments for measuring evaluation capacity are also lacking evidence of recommended validation processes and practices. Validation of EC instruments is most often not carried out, and in those cases where some validity evidence is collected, the validation is not explicitly described and discussed (Cousins et al., 2008; Nielsen et al., 2011; Taylor-Ritzler et al., 2013). Moreover, the context in which the instrument is used is usually not considered (Suarez-Balcazar et al., 2010), and even when it is considered, either it is not incorporated during the validation process or there is no explicit description of that process.

In this study, validity evidence was gathered beginning with the development of the EC instrument using multiple sources of data to capture the context (i.e., the early childhood field), providing potential users with an understanding of the appropriateness of the instrument so that they can make informed decisions about using it. To improve the validity of the EC measurement, an argument-based approach to validity was adopted, which is discussed in the next section.

The Argument-Based Approach to Validity

In this study, validity is established through an argument-based approach in which empirical evidence is provided to substantiate or refute inferences made from the data generated through use of an instrument (Kane, 2013a). Consistent with the argument-based view, the Standards for Educational and Psychological Testing (henceforth *the Standards*; AERA et al., 2014) emphasize the importance of evaluating the plausibility of an interpretation of instrument scores (i.e., the data; Kane, 2013a). From this perspective, validity is not considered as a singular state to be achieved, but rather involves a dynamic process in which evidence is gathered from multiple sources of data (AERA et al., 2014). Thus, the present study aligns with the 1999 and 2014 standards, in which establishing validity involves generating supportive evidence for the inferences drawn from the data for a specific sample.

This approach to gathering the evidence for validity, which is called *validation*, is "an ongoing process in which various sources of validity evidence are accumulated and synthesized to support the construct validity of the interpretation and use of instruments" (Hubley & Zumbo, 2013, p. 4). Validation is the process by which the interpretation and use of instrument scores and any underlining assumptions are evaluated for coherence and completeness (Kane, 2013a). Validation thus begins with an explicit definition of the construct and a statement of the

proposed interpretation and use of the scores (Kane, 2013a). It continues throughout the development, administration, and use of the instrument (Hubley & Zumbo, 2011; Kane, 2013a).

The argument-based approach to validity is one of the contemporary views that have replaced earlier formulations in which validity was considered a property of an instrument (AERA et al., 2014; Sireci, 2009), implying that validity is a stable characteristic of the instrument (Gunnell et al., 2014). The contemporary view, in contrast, views validity as a dynamic property of the scores obtained from the instrument in the context in which validity is being examined (e.g., AERA et al., 2014; Gunnell et al., 2014). As a dynamic concept, validity is not judged dichotomously in terms of being either valid or invalid, as in earlier views, but instead, it is seen as a matter of degree (Hubley & Zumbo, 2013). Validity is therefore not absolute; rather, it involves making the most sensible justification guiding the use of the instrument and research to advance our understanding of the results (Messick, 1989).

To facilitate the application of the unified view, I examine five sources of validity evidence that support the validation process. Given that the unified view is an argument-based view, obtaining multiple sources of evidence that can be used to substantiate claims about the validity of inferences leads to greater confidence in the overall conclusion (Kane, 2013a). The five sources of evidence are: content, the response process, internal structure, relationship to other variables, and consequences of testing (AERA et al., 2014). The first source of evidence, content evidence, focuses on the instrument's wording, its themes, and item types (Gunnell et al., 2014). The content evidence that is gathered can include expert reviews, theoretical support, and quantitative data (AERA et al., 2014). The second source of evidence, the response process, provides evidence based on the cognitive processes involved when respondents answer the instrument's items (Chan, 2014). Techniques such as having respondents think aloud, cognitive

interviews, and expert reviews can be used to collect response process evidence (AERA et al., 2014).

The third source of evidence, the internal structure of the instrument, includes the dimensionality or the factor structure of the instrument (Sireci & Sukin, 2013). Specifically, the internal structure evidence reveals "the degree to which the relationships among test items and test components conform to the construct on which the proposed test score interpretations are based" (AERA et al., 2014, p. 16). Statistical techniques such as factor analysis, structural equation modeling, and item response theory can be used to generate supportive evidence (AERA et al., 2014; Kane, 2013a). The fourth source of evidence, relations to other variables, includes the relationship of the instrument scores for subconstructs to one another and to other external variables measured by a validated instrument with related or different constructs (AERA et al., 2014). Evidence for this source can include convergent, discriminate, and predictive criteria (AERA et al., 2014).

Finally, the fifth source of evidence, the consequences of testing, includes the intended and unintended as well as positive and negative outcomes of using an instrument (AERA et al., 2014; Sireci & Sukin, 2013). Supportive evidence for consequences can be gathered by explicitly discussing the intended and unintended use and potential consequences and by consulting experts in the field for which the instrument is intended (AERA et al., 2014). Given that validity is about making the most sensible and coherent judgment about inferences, such judgment is enhanced by providing multiple sources of evidence and considering the context and the sample for the use of the instrument (Hubley & Zumbo, 2011).

Towards Valid Evaluation Capacity Instrumentation

The multifaceted nature of EC creates challenging conditions for researchers seeking to define, conceptualize, and measure it (Cousins, Goh, Elliott, Aubry, & Gilbert, 2014). Conducting EC research is further complicated by the context dependence of EC, which means that the results are specific to the context in which they were studied and may have limited transferability across contexts. In the following subsections, I summarize the gaps in the current EC literature related to the lack of (a) theoretical foundations grounding EC measurement, (b) context specificity, and (c) an integrative approach to creating EC instruments. Each of these subsections explains the importance of the given topic, synthesizes the current literature focused on the topic, delineates the limitation, and finally advances the new approach adopted in this study.

Theoretical foundation grounding evaluation capacity measurement. Theories are a fundamental aspect of instrument development because they allow for a better understanding of the construct(s) to be measured and the validation approach that is required (Hubley & Zumbo, 2011; Kane, 2013a). Measurement involves a set of rules that delineate how values are assigned to objects, behaviours, or actions (Gamst, Meyers, & Guarino, 2008). These rules are critical for establishing confidence in the inferences made by a measurement instrument (Raykov & Marcoulides, 2011). Given that the rules are not prescriptive and that the researcher needs to use his or her judgment in developing and validating an instrument, a clear process through which decisions are made is required (Raykov & Marcoulides, 2011). Articulation of a measurement process is rooted in validity theory (Kane, 2013a). For example, the process may involve specifying the validation approach that will be followed, the operationalization of the construct to be measured, and the relationship between the items and scales comprising the instrument.

This articulation may help users to evaluate the appropriateness of the instrument for meeting their needs and to assess claims about the validity of the inferences based on the data (Shepard, 2016). The current EC literature lacks a clear link between the theoretical conceptualizations of the EC construct and the instruments that have been developed to measure EC. Furthermore, the literature describing the EC instruments that have been developed lacks a clear measurement process. The lack of theoretical link and guiding literature represent two limitations that weaken the validity and reliability of the inferences made from the data generated by current EC instruments and consequently their usefulness to others.

The first limitation is an absence of a clear link between the conceptualizations of EC or the EC building construct and the instruments designed to measure EC. This impairs the validity and reliability of inferences about EC that are reached using an instrument (Kane, 2013a). Several definitions of EC and EC building have been put forth in the literature (e.g., Boyle & Lemaire, 1999; King & Volkov, 2005; Preskill & Boyle, 2008a; Stockdill, Baizerman, & Compton, 2002), but with little agreement (Labin et al., 2012). Similarly, although the EC literature provides frameworks that conceptualize this construct (Cousins et al., 2008; King & Volkov, 2005; Preskill & Boyle, 2008a), they vary in the details they provide about the components, how the components are defined, and the relationships among the components. Operationalizing a latent construct in order to measure it is important because this provides a foundation for conceptualizing it, defining the components (i.e., factors) involved in the conceptualization, and creating a pool of items to measure each of the components (Raykov & Marcoulides, 2011). The measurement occurs indirectly through proxies or indicators—that is, observed manifestations of behaviours that are believed to relate to the latent construct (Raykov & Marcoulides, 2011). Assignment of such proxies or indicators must be supported by a

rationale based on a theoretical understanding, empirical research, practice, or some combination of these (Kane, 2013b; Zumbo, 2009).

The studies that have proposed existing EC instruments do not establish a clear connection between their conceptualization of the construct and its measurement (for further background on the developmental procedures see Nielsen et al., 2011; Preskill & Torres, 2001; Taylor-Ritzler et al., 2013). To varying degrees, the authors of EC instruments such as the Evaluation Capacity Assessment Instrument (ECAI; Taylor-Ritzler et al., 2013), the Evaluation Capacity Index (ECI; Nielsen et al., 2011), the Readiness for Organizational Learning and Evaluation tool (Role; Preskill & Torres, 2001), and the Evaluation and Organizational Capacity Survey (Cousins et al., 2008) fall short in establishing a clear link between the instrument and the theoretical rationale used to define the construct and its components.

A review of evaluation instruments reveals some missing links that limit our ability to have confidence in the instrument findings. For example, Taylor-Ritzler and colleagues (2013) adopted an EC framework developed by Suarez-Balcazar and colleagues (2010), but they did not explain the relationship between the instrument and the framework, and they also failed to provide a rationale for the changes they made to the framework. The lack of an articulation of the relationship and rationale makes it difficult to draw clear inferences from the data, which weakens the validity and reliability of the inferences. Similarly, Cousins and colleagues (2008) developed an EC framework but did not explain the relationship between their survey items and the components of the framework, nor did they explain how they selected the items to define the constructs. The lack of content clarity in these instruments is problematic for establishing validity evidence.

The second limitation mentioned is that the existing EC literature does not ground instrument development in validity theory and lacks a clear rationale for decisions made throughout the validation process (Cheng & King, 2016; Labin, 2014). Most of the existing instruments have not been subjected to validation (Labin et al., 2012; Taylor-Ritzler et al., 2013). Instrument validation involves providing clear procedures for gathering validity evidence that can be used to substantiate the proposed inferences (AERA et al., 2014). The validation process can therefore help researchers and potential users to evaluate an instrument and the inferences generated from the data (AERA et al., 1999, 2014).

Context specificity. Creating an instrument that is contextually relevant requires a process in which the influences of the context are carefully considered (King & Volkov, 2005; Zumbo, 2009). Context specificity means establishing the context in which the instrument will be used (Hubley & Zumbo, 2011; Zumbo, 2009). Context is important for EC measurement in terms of instrument validation (Hubley & Zumbo, 2011; Zumbo, 2009) and capturing the complexity of building EC (King & Volkov, 2005).

Contextualizing measurement has been a well-established practice for decades in other fields of research, including educational measurement and psychological testing (Kane, 2013b; Messick, 1989; Shepard, 2016). In the 1970s, for example, several court cases attested to the significance of context in interpreting measures (Shepard, 2016). For example, in Griggs v. Duke Power Co. in 1971, the United States Supreme Court "ruled against the use of intelligence tests to select employees for higher level jobs because the test had a discriminatory impact on blacks, and the company lacked evidence that the test was related to job performance" (Shepard, 2016, p. 271). In measurement theory, an instrument is created to address a need of its specific users, and the validation of the claims made and their consequences cannot be separated from the

context for which the instrument was created (Shepard, 2016; Zumbo, 2009). Therefore, validation of the claims generated by an EC instrument must begin during its development process, specifically, during the conceptualization of the construct being measured (Kane, 2013a; Zumbo, 2009). A failure to explicitly consider context will inevitably hinder interpretations and weaken the claims put forward about the validity of the inferences.

More recently, in the field of EC building, King (2007, 2017) has argued that understanding context is crucial for building capacity. Other researchers have similarly argued that context is particularly important in building EC because both the organization and the community it serves exist within a social, historical, economic, and political milieu (e.g., SenGupta, Hopson, & Thompson-Robinson, 2004; Suarez-Balcazar et al., 2010). Although there is agreement on the centrality of context for building EC, much of the current literature fails to consider the importance of context in conceptualizing and measuring the EC construct (Suarez-Balcazar et al., 2010). Failure to consider context may limit the accuracy of the data generated by the instrument, which may result in inaccurate decisions based on incomplete information. This issue manifests in the literature representing two limitations: a failure to consider the context in which the instrument will be implemented, and a failure to provide validity evidence when the context is implicitly mentioned.

The initial limitation involves a failure to ground the theoretical and empirical understanding of the construct within a specific context to address the conceptualization and measurement of EC (e.g., King and Volkov, 2005; Preskill & Torres, 2001; Taylor-Powell & Boyd, 2008; Taylor-Ritzler et al., 2013). Specifying the context is especially important in conceptualizing EC because it is a multifaceted construct that is not yet clearly understood empirically (Cheng & King, 2016; Gagnon, Aubry, Cousins, Goh, & Elliott, 2018; Labin, 2014).

The lack of such understanding is apparent in existing theoretical conceptualizations; although they reflect some agreement on key components that impact EC building (e.g., organizational learning, organizational leadership, individual attitudes regarding evaluation), they are still inconsistent with respect to how they define these components, their relationship to EC, and their relationships to each other (King, in press; Nielsen et al., 2011; Taylor-Ritzler et al., 2013). For example, Preskill and Torres (2001) and Preskill and Boyle (2008a) successively developed an instrument (The Readiness for Organizational Learning and Evaluation Instrument) and a framework (Multidisciplinary Model of Evaluation Capacity Building) that they suggest are applicable "to all ECB [evaluation capacity building] contexts" (Preskill & Boyle, 2008a, p. 444). Their claim might be true; however, an investigation of the claim is warranted that provides validity evidence for each context in which the instrument is used.

The second limitation applies to studies in which the authors make a passing reference to context but do not claim context relevance (e.g., King & Volkov, 2005; Nielsen et al., 2011). In these studies, the authors fail to provide evidence supporting context relevance and the ways in which the context has influenced the inferences that are made. This is problematic because it weakens the validity of inferences generated from instruments that are not specific to a context (Zumbo, 2009). It is crucial to be clear from the outset about the context in which the instrument will be administered, how the instrument's purpose (s) is (are) relevant to the context, and the potential consequences of its use in the specified context (Zumbo, 2009).

Considering the context is essential in building and studying EC (Suarez-Balcazar et al., 2010) because this allows researchers to create an EC conceptualization that is relevant to the given context, which in turn informs the development of an instrument that may be used to empirically assess this conceptualization. Creating an instrument or a framework that is

contextually relevant requires an approach to development that embeds ways to capture context throughout the entire process (Zumbo, 2009). Such an approach may include involving diverse stakeholders in the operationalization of the construct and the items (Luyt, 2012). It may also require administering the instrument either in a specific context or in multiple contexts. If the expected sample size is large, this would allow for statistical tests comparing respondents in different contexts in order to examine measurement invariance between the different groups (AERA et al., 2014).

Integrative methodological approach. Choosing a research method requires making choices, which involve trade-offs (Visser, Krosnick, & Lavarakas, 2000). For example, qualitative methods provide deeper understanding but less generalizability than can be achieved through quantitative methods. A research methodology is a plan of action that incorporates methods or techniques and procedures used to gather, analyze, and interpret data (Creswell & Plano-Clark, 2011). Many methodologists have advocated using multiple methods to triangulate across methods in order to minimize the weaknesses and maximize the strengths associated with each method taken alone (Campbell & Fiske, 1959; Creswell, 2009). Triangulation can involve different types of data, both qualitative and quantitative, and different techniques for each type (Greene, 2007). Triangulation improves rigour in the cases of convergence and divergence of findings (Greene, 2007). While the importance of rigour is well established in measurement theory, EC studies that focus on developing frameworks and instruments have not always achieved rigour. Therefore, two limitation are apparent in EC studies: 1) the procedure used during EC instrument development and/or validation of the developed instruments; and 2) lack of triangulation of different data sources.

Weaknesses in procedure. The first limitation involves use of weak procedures in developing EC instruments. Survey design should involve procedures that are transparent (Wolf, Joye, Smith, & Fu, 2016). A survey or an instrument is "a systematic method for gathering information from (a sample of) entities for the purpose of constructing quantitative descriptors of the attributes of the large population of which the entities are members" (Groves, Fowler, Couper, Lepkowski, & Singer, 2004, p. 2). The transparency of the procedures and the type and size of the sample used in the development and validation are critical in establishing the rigour of the methodological approach, which in turn enhances the usefulness and accuracy of the resulting inferences (Blair, Czaja, & Blair, 2013).

The current EC literature provides insufficient descriptions of the procedures followed during the development and validation of EC instruments. Having a clear procedure allows other researchers to evaluate the methodology and better understand the results (Kane, 2013b). A distinct procedure also provides researchers with a way to replicate the study, which can advance our understanding of EC using a different sample that may potentially increase the generalizability of the findings in quantitative studies and the richness of what we know through qualitative studies. This lack of procedural clarity manifests in studies where the authors have consulted experts to evaluate an instrument but do not provide a description of the process that was used, such as eliciting experts' recommendations and what those recommendations were. Articulating the process allows other researchers to evaluate it and address any limitations in future studies. The importance of this transparency is exemplified by Taylor-Ritzler and colleagues' (2013) study, in which they consulted three evaluation experts to review the instrument they developed but do not appear to have consulted an expert in measurement or context to evaluate the appropriateness of the scales or the items.

The nature of the sampling and the type and size of the sample for existing EC instruments also present limitations that weaken the validity of the inferences that are made and limit their generalizability. Researchers have used purposeful instead of random sampling because the organizations that conduct evaluation and may require EC are not well defined (e.g., Taylor-Ritzler et al., 2013). In quantitative research, the generalizability of inferences requires that the sample represents the population of interest (Tashakkori & Teddlie, 2003; Zumbo, 2009). As an example of these limitations, the samples used by Taylor-Ritzler et al. (2013) were intended to represent the not-for-profit organizations in the Chicago area. However, the sample largely consisted of executive directors whose knowledge of evaluation might not be representative of all employees, such as front-line workers (Taylor-Ritzler et al., 2013). Such a methodological limitation, which is acknowledged by the authors themselves, limits the possible inferences that can be made. For instance, Fierro (2017) compared managers' and front-line employees' perceptions of existing EC and found that the managers tended to report a more positive view of EC than did the front-line employees. Therefore, having uneven representations may compromise the validity of the inferences that are made using data from the instrument. In addition, a small sample combined with many indicators and the breath and diversity of the sample limit the conclusions that can be drawn from the analysis (e.g., Cousins et al., 2008; Taylor-Ritzler et al., 2013). Our understanding of conceptualizing and measuring EC will be expanded by addressing several of these issues.

Lack of triangulation. A second limitation of the EC literature is the frequent use of a single methodological approach during the development or validation of EC instruments (e.g., Cousins et al., 2008; King & Volkov, 2007; Preskill & Boyle, 2008b). The particular evidence provided in these studies is largely descriptive (e.g., Cousins et al., 2008; Fierro, 2012).

Reliability analysis using the coefficient alpha (Cronbach, 1951) and descriptive statistics are among the main statistics that are usually provided (e.g., Cousins et al., 2008; Fierro, 2012). More complex analyses such as factor analysis and structural equation modeling have been used to gather validity evidence for a smaller number of instruments (for more details on their statistical analysis see Nielsen et al., 2011; Taylor-Ritzler et al., 2013). While a quantitative approach is valuable for investigating relationships among variables, testing the conceptual frameworks followed, and providing generalizable findings, maximization of its contributions requires grounding the indicators in theory and context (AERA et al., 2014; Creswell, 2009). In fact, the unified view of validity recommends the use of multiple sources of validity evidence to strengthen validity arguments (AERA et al., 2014). Using multiple data sources and statistical analyses in combination will strengthen the validity of the claims made based on an instrument.

To address these challenges in EC measurement and conceptualization, this study uses a multimethod approach, guided by the principles of the unified view of validity, in the development and validation of an EC instrument that is grounded in theoretical and contextual understandings. The next section begins the theoretical understanding of EC by synthesizing the literature in terms of existing EC conceptualizations and measurements of EC components. The synthesis provides a starting point for conceptualizing EC and developing an EC instrument in ways that are theoretically sound and contextually relevant to the early childhood field.

Conceptualizing and Measuring Evaluation Capacity

To develop a new EC instrument that is grounded in theory, it is necessary to survey the components measured in existing instruments and analyze their operationalization. The focus of this study is on operationalizing individual and organizational components that are directly related to EC or EC building. The numbers and types of EC components reported in these

instruments vary widely; however, researchers report commonalities across elements that are widely believed to be crucial contributors to EC (Bourgeois, Chouinard, & Cousins, 2008; Cheng & King, 2016; Labin, 2014). Identifying the common features across seven relevant instruments, summarized in Table 1, helped to inform the conceptualization of EC and the development of the EC instrument.

Comparisons across the instruments in terms of formats and measured constructs reveal the frequent use of the survey format with a dual focus on conducting and using evaluations. It is interesting to note that only four of the instruments in the table include items measuring EC at both the individual and organizational levels, with the remaining three instruments focusing only on the organizational level. One of the greatest limitations of the instruments is the lack of an indication in their descriptions that all five sources of validity evidence (content, internal structure, response process, consequences, and relation to other variables) have been addressed. The most frequent type of validity evidence that is reported is related to content; by contrast, fewer than half of the instruments have described validity evidence related to internal structure. Finally, an inspection of the components comprising individual and organizational EC reveals some commonalities as well as some notable differences, which are explained in the following section. Examining the common features across these instruments provides an essential foundation for the study's development of a survey instrument with items measuring both individual and organizational levels and the sources of validity evidence generated.

Table 1.

Relevant Evaluation Capacity Building Instruments Informed this Study

Overall Themes *Common Features*	Volkov and King (2007)	Cousins et al. (2008)	Nielsen et al. (2011)	Instruments Fierro (2012)	Bourgeois and Cousins (2013)	Taylor-Ritzler et al. (2013)	Cheng and King (2016)
Instrument format							
Survey	--	✓	✓	✓	--	✓	✓
Checklist	✓	--	--	--	--	--	--
Rubric	--	--	--	--	✓	--	--
Measured constructs							
Conduct evaluation	✓	✓	✓	✓	✓	✓	✓
Use evaluation	--	✓	✓	✓	✓	✓	--
Level of focus							
Individual	--	--	✓	✓	--	✓	✓
Organizational	✓	✓	✓	✓	✓	✓	✓
Sources of validity							
Content	--	✓	✓	✓	✓	✓	✓
Internal structure	--	--	✓	--	--	✓	--
Response process	--	--	--	--	--	--	✓
Individual components							
Behavioural beliefs	--	--	--	✓	--	✓	✓
Behavioural controls	--	--	✓	✓	✓	✓	✓
Behavioural goals	--	--	--	--	--	✓	✓
Organizational components							
Behavioural beliefs	✓	✓	✓	✓	--	✓	✓
Behavioural controls	✓	✓	✓	✓	✓	✓	✓
Behavioural goals	--	✓	--	✓	✓	✓	✓

Note. ✓: Feature addressed by instrument. --: Feature not addressed by the instrument

This section is organized in two parts, one addressing individual EC and the other organizational EC. To understand the operationalization of these components, I turn to the theory of planned behaviour (Ajzen, 2005; Ajzen & Fishbein, 1977; Fishbein & Ajzen, 2011), which offers a lens for identifying shared understanding. Consistent with this theory, my analysis of the components of individual and organizational EC addresses three aspects of behaviour: behavioural beliefs, behavioural controls, and behavioural goals. *Behavioural beliefs* are stances that may be positive or negative regarding performing a specific behaviour (e.g., collecting data for evaluation purposes). In large part, a person's positive perception of the outcome resulting from performing that behaviour (e.g., believing that the data will improve the decisions that are made) is likely to result in a positive adoption of that behaviour (Fishbein & Ajzen, 2011). *Behavioural controls* are elements (e.g., interview skills for data collection or resources for conducting interviews) that are required to perform the behaviour (Fishbein & Ajzen, 2011). The third aspect of behaviour, *behavioural goals*, refers to the end that is sought (e.g., making better decisions), based on one or more preceding actions (Fishbein & Ajzen, 2011). I first explain each aspect of behaviour and then highlight the EC components exemplifying that aspect. These insights add to our understanding and create takeaways for advancing the conceptualization and measurement of EC. They also play a key role in informing the development of this study's hypothesized EC framework and instrument presented in Chapter 3.

Individual evaluation capacity components. Examination of the individual EC components in terms of the three aspects of behaviour reveals that although all three aspects are represented in the instruments, the instruments tend to focus on behavioural controls more than beliefs and goals. In addition, only one instrument by Taylor-Ritzler and colleagues (2013) appears to address all three aspects of behaviour.

Behavioural beliefs. Individual EC components addressing behavioural beliefs appear in three of the EC instruments in Table 1 (i.e., Fierro, 2012; Taylor-Ritzler et al., 2013; Cheng & King, 2016). At the individual level, behavioural beliefs concern an individual's attitude toward performing a behaviour (Fishbein & Ajzen, 2011). Only two instruments included components measuring this aspect of behaviour, labeled as *Awareness* (Taylor-Ritzler et al., 2013) and *Attitude* (Fierro, 2012). Both operationalize this aspect of behaviour in terms of respondents' perceptions of the usefulness of evaluation. For example, Taylor-Ritzler and colleagues used items measuring respondents' beliefs about the usefulness of evaluation to their work and the possible purposes it might help achieve, such as making decisions, improving a program, or justifying funding. Fierro (2012) used items that addressed similar points to those in Taylor-Ritzler and colleagues' instrument, but were worded in more general ways (e.g., "Evaluation has the potential to add value to the work we do," p. 204).

The fact that only two instruments included items measuring attitudes or beliefs is not reflective of their importance in the literature conceptualizing EC or EC building. Scholars have reported that individual's attitudes or beliefs about evaluation may influence whether they seek to improve their ability to conduct and use evaluations, such as by attending relevant training sessions (Fleming, 2011). For example, in their systematic review, Labin and colleagues (2012) found that organizations focused little of their attention on targeting employees' attitudes and that a negative attitude was frequently cited as a hurdle to undertaking and using evaluations. The authors concluded that there is a need for efforts to improve organizational members' attitudes toward evaluation.

Behavioural controls. All five instruments that measured individual EC included items that correspond to behavioural controls. For individual EC, this aspect of behaviour refers to the

elements that are required to make it possible for an individual to act on an intention to perform the behaviour (Fishbein & Ajzen, 2011). Perceiving a sense of control over the desired behaviour provides the individual with a sense of self-efficacy—that is, the "person's estimate of his or her capacity to orchestrate performance on a specific task" (Gist & Mitchell, 1992, p. 183). For example, lacking the evaluation skills required to conduct aspects of evaluation (e.g., analyzing qualitative data) hinders an individual's ability to participate in at least some aspects of evaluation. Instruments, including those focusing on EC, can only measure a person's perceived control over performing a task, which serves as a proxy for the actual control over completing the identified task (e.g., conducting and using evaluations). In the surveyed EC instruments, components measuring behavioural controls have singularly focused on measuring evaluation knowledge and skills.

The evaluation knowledge and skills component of EC has long been a focus of the EC literature (Labin et al., 2012). Overall, instruments measuring EC tend to focus on the wide range of knowledge and skills pertinent to evaluation, which are reported in the Canadian Evaluation Standards (Yarbrough, Shulha, Hopson, & Caruthers, 2011). In all instruments that are reviewed in this section, evaluation skills include technical practices (e.g., evaluation planning, data collection, data analysis, interpretation, and writing) and interpersonal practices (e.g., communication, collaboration).

Behavioural goals. Only one instrument in Table 1 (Taylor-Ritzler et al., 2013) includes a component, Motivation to Engage in Evaluation that exemplifies behavioural goals – which are an individual's intended purposes or actions. While an immediate desired action in EC is to engage in conducting and using evaluations, an instrument can only measure a proxy of this goal, which is related to the person's motivation (Ajzen, 2005). According to research in psychology,

a person's motivation can be measured by the "the degree to which a goal-relevant object is evaluated positively" (Touré-Tillery & Fishbach, 2014, p. 330). Making this explicit may involve inquiring about the person's willingness to participate in a specific task (e.g., an aspect of evaluation). Among the instruments reviewed in this section, Taylor-Ritzler and colleagues (2013) took a more direct approach, using the term *motivation* in a question asking about the respondent's motivations to learn about evaluation, evaluate their programs, and support their staff in evaluating programs.

The relative absence of the motivation component in existing EC instruments is not reflective of its importance in the EC literature focused on conceptualizing EC. According to the EC literature, this component is key for influencing employees' involvement in evaluation activities and using evaluation results (e.g., Cheng & King, 2016; Preskill & Boyle, 2008). Employees who are motivated to conduct and use evaluations are likely to develop their EC (e.g., knowledge and skills) and to show an interest in being involved in evaluation processes and using the resulting conclusions/findings (Burke, Lake, & Paine, 2008; Clinton, 2014; Labin, 2014). In practice, however, there is a disconnect (Labin et al., 2012). While Labin et al.'s (2012) literature synthesis emphasizes the importance of motivation in developing EC, organizations have rarely reported that their EC efforts targeted motivation, even when lack of motivation was thought to be a hindrance to building EC. Being clear about what the immediate goal of EC is appears to be the first step in achieving it.

Organizational evaluation capacity components. Similar to the individual EC components, examination of the organizational EC components in terms of the three aspects of behaviour reveals that behavioural controls tend to be the most represented, followed by

behavioural beliefs. Furthermore, three instruments (Cousins et al., 2008; Fierro, 2012; Taylor-Ritzler et al., 2013) appear to address all three aspects of behaviour.

Behavioural beliefs. Organizational EC components that clearly address behavioural beliefs appear in five of the seven instruments in Table 1 (Cheng & King, 2016; Cousins et al., 2008; Fierro, 2012; Taylor-Ritzler et al., 2013; Volkov & King, 2007). For organizations, behavioural beliefs refer to norms that reflect the typical established attitudes or views toward a behaviour, such as conducting an evaluation. These normative beliefs tend to influence the "perceived social pressure to engage or not in the behaviour" (Fishbein & Ajzen, 2011, p. 20). The organizational EC components in the instruments related to this aspect of behaviour include Organizational Context (Volkov & King, 2007), Learning Climate (Taylor-Ritzler et al., 2013), Evaluation Culture (Cheng & King, 2016), and Leadership (Bourgeois & Cousins, 2013; Taylor-Ritzler et al., 2013). These components share a common thread, one that may signal that an organization is committed to conducting and using evaluations. Few of the existing instruments focus on characteristics that allow evaluation practices to grow, such as openness to new ideas and risk-taking (e.g., Taylor-Ritzler et al., 2013). Some instruments focus on perceived evidence of organizational practices, including those of its leadership, that may reflect a commitment to conducting and using evaluations, such as using information learned from evaluations to improve organizational processes and decision-making (Cheng & King, 2007; Taylor-Ritzler et al., 2013). Other instruments focus on whether organizations have policies and procedures that may reflect an organizational commitment to evaluation, such as expecting new employees to have an orientation toward conducting evaluations (Bourgeois & Cousins, 2013; Fierro, 2012).

The importance of organizational norms is well established in the EC literature. Scholars have discussed the importance of an evaluation culture as indicative of the degree to which EC

efforts may be successful (Labin et al., 2012; McCoy, Rose, & Connolly, 2013; Suarez-Balcazar et al., 2010). There seems to be a consensus that a positive evaluation culture or environment is distinguished by regular use of evaluation practices such as planning, data collection, and disseminating and using results to inform program improvement (Botcheva, White, & Huffman, 2002; Cheng & King, 2016; McCoy et al., 2013; Preskill & Boyle, 2008a). Another factor that influences and is influenced by the established culture is organizational leadership (e.g., Burke et al., 2008; Burke & Litwin, 1992). Organizational leaders who view evaluation positively are likely to adopt policies and practices that are supportive of evaluation, such as investing resources to support their employees in conducting and using evaluations (Cousins, Bourgeois, & Associates, 2014; Milstein, Chapel, Wetterhall, & Cotton, 2002; Preskill & Boyle, 2008b). Organizational leaders are also influential in the extent to which information generated from evaluations is used, and this can be reflected in the regular discussion of use of their findings in staff meetings (King & Volkov, 2005). In contrast, leaders with negative views of evaluation will tend to negatively influence their employees' sense of behavioural control, making it difficult to conduct and use evaluations.

   Behavioural controls. All seven instruments include components that exemplify this aspect of behaviour. Behavioural controls are necessary resources provided by organizations to make it possible to perform a behaviour (Fishbein & Ajzen, 2011). While behavioural controls at the individual level are within an individual's own control, at the organizational level, they involve those elements that impact a person's ability to perform a behaviour, such as conducting an evaluation. For example, it is challenging for individuals to make time for evaluation activities if the organization does not explicitly make this one of the key tasks that employees are expected to do. In such cases, an individual may have the necessary skills and attitudes to

conduct and use evaluations; but lacking the needed resources such as time, money, and support

for data collection hinders that individual's ability to conduct evaluations. The organizational

components in this category measured by the existing instruments include Organizational

Support Structures (Cousins et al., 2008), Resources (Bourgeois & Cousins, 2013; Fierro, 2012;

Volkov & King, 2007), and access to information about evaluation (Fierro, 2012).

While the importance of components that exemplify behavioural controls is not in

dispute, they are often among the challenges that organizations must deal with when carrying out

evaluations (Labin et al., 2012). Sometimes, this is understandably so because funding,

especially for not-for-profit or community organizations, is very limited (Gokiert et al., 2017;

Nakaima & Sridharan, in press; Runnels, Andrew, & Rae, 2017). If evaluation plays an

important role in an organization for enhancing the quality of their programs and decision-

making, this importance should be reflected in the organizational structures and the availability

of the resources needed for evaluation (Patton, 2011a).

Behavioural goals. Four of the reviewed instruments include components that reflect

behavioural goals (see Table 1; Bourgeois & Cousins, 2013; Cousins et al., 2008; Fierro, 2012;

Taylor-Ritzler et al., 2013). At the organizational level, behavioural goals are those sought by the

organization, which should be aligned with individuals' goals. The components in this category

are for the most part similar in that they explicitly or implicitly measure Mainstreaming Use of

Evaluation (also labelled Integration with Organizational Decision-Making [Fierro, 2012], and

Use of Evaluation Process [Cousins et al., 2008]). Mainstreaming Use of Evaluation refers to

"the process of making evaluation an integral part of an organization's everyday operations"

(Sanders, 2003, p. 3). The extent to which evaluation processes and use of the results are an

integral aspect of an organization may be reflective of its evaluation culture (Duignan, 2003;

Fierro, 2012; Taylor-Ritzler et al., 2013). Scholars who have operationalized mainstreaming have generally focused on the extent to which evaluation is embedded in different activities over time and the breadth of those who are involved in evaluation (Bourgeois & Cousins, 2013; Cousins et al., 2008; Fierro, 2012; Taylor-Ritzler et al., 2013). To assess the extent to which evaluation is embedded within various activities, scholars have used items asking how often program staff ask about evaluation and how often evaluation has been part of an agenda (Fierro, 2012).

The EC literature appears to agree that making evaluation activities and use an integral part of the organization is crucial because of the positive benefits evaluations have for delivering useful programs and services (Patton, 2011a; Wilcox & King, 2013). Organizational Readiness for Evaluation, which is discussed in the literature conceptualizing EC (Walker-Egea, 2014) but was not measured by the EC instruments reviewed in this section, is similar to Mainstreaming Use of Evaluation. Organizational Readiness for Evaluation and Mainstreaming are complex constructs and share similarities with the evaluation culture and evaluative thinking (King, in press). Arguably, given this complexity, these components may be multidimensional constructs that require their own conceptualization and instruments for measurement.

Study Need and Research Questions

An empirical conceptualization and measurement of EC reflective of the contemporary views of validation remains a gap in the literature that this study begins to address. Organizations are increasingly being required to conduct evaluations that meet diverse stakeholder needs, and this demand for evaluation requires them to build EC (e.g., Huffman, Thomas, & Lawrenz, 2008; Preskill & Boyle, 2008b). The complexity inherent in EC building necessitates a better and more empirical understanding of the construct itself and the process for building EC in practice (Cousins et al., 2014; Labin, 2014; Nielsen et al., 2011). Practitioners need to know what types of capacity to develop, how to develop it, and what outcomes to expect from such efforts, given their specific context (King & Volkov, 2005). Answering these questions requires an empirical investigation that builds on existing theoretical and empirical knowledge and integrates the areas of evaluation, measurement, and methodology. Existing literature has advanced our knowledge of EC, especially regarding its definition and conceptualization; however, several limitations still need to be addressed. Aiming to address these, the present study was guided by the following research questions:

1. What are the underlying EC factors in the early childhood field?

2. To what extent do the sample data support the theorized EC model in the early childhood development field?

3. Do professional roles influence any of the paths connecting the individual and organizational EC constructs? If so, how?

Chapter 3: Methodology

The purpose of this study was to develop and validate a contextually relevant instrument for measuring individual and organizational EC in early childhood development organizations in Alberta, Canada. This was accomplished by following procedures for developing an EC instrument, collecting data within the early childhood field using the instrument, and analyzing the psychometric properties of the instrument. These procedures generated four types of validity evidence (content, internal structure, consequences, and relations to other variables) according to five measurement standards (AERA et al., 2014). This chapter describes the methods and procedures used to develop and validate the EC instrument. It begins with a presentation of the study's research design and ethical considerations; then it turns to the procedures involved in the instrument development, data collection, and data analysis.

Research Design

A cross-sectional survey design (Creswell, 2009) was used to answer the research questions. First, the items for the instrument were developed based on the literature review and expert consultations, and expert reviews were then conducted to evaluate the instrument and assess its contextual relevance. Next, data were collected from participants in the early childhood field in Alberta, Canada. Finally, a series of data analyses were conducted to assess the psychometric properties of the instrument. This process generated validity evidence specific to the instrument's content and consequences from the expert reviews as well as validity evidence related to internal consistency from the psychometric analyses. Figure 1 provides an overview of the development and validation process undertaken in this study.
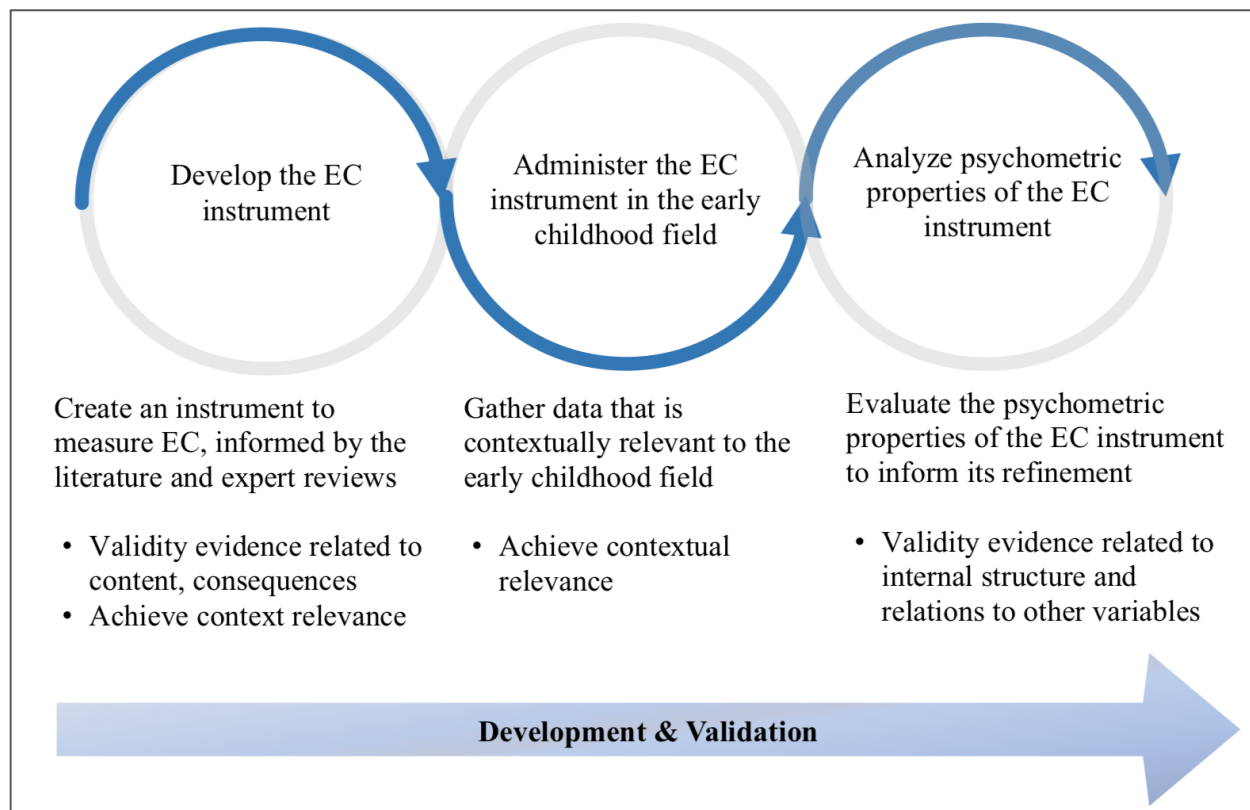
Figure 1. Evaluation capacity instrument development and validation process.

Five measurement standards were adopted for this study to provide guidance during each aspect of the instrument development and validation (AERA et al., 2014). Table 2 summarizes the focus of the five standards—1.1, 1.11, 1.13, 1.16, and 2.3—and the related strategies used to gather validity evidence in this study. Standards 1.1 and 1.11 emphasize the need to foresee potential uses of the instrument and interpretations during the development of the instrument. I addressed these two standards during the instrument development by (a) consulting the overall project committees during the conceptualization and operationalization processes and (b) conducting expert reviews to evaluate the instrument in terms of content, measurement, and appropriateness for the early childhood context. Standards 1.13 and 1.16 stress the importance of

gathering evidence related to the internal structure of the instrument and the relationships among related constructs. I addressed these two standards during data analysis by conducting confirmatory factor analysis (CFA), structural equation modeling (SEM), and multi-group path analysis to investigate the relationships among individual and organizational factors and the influence of respondents' roles on these relationships. Standard 2.3 focuses on the reliability of the factors, which I investigated by evaluating the coefficient alpha values for the EC scales. My approach in this study conforms to the contemporary view of validity (AERA et al., 2014), in which multiple types of evidence are produced during the validation process, beginning with the development of the instrument and continuing thereafter to substantiate the validity argument.

Table 2.

Key Validation Standards Guiding Development of Study Instrument

AERA, APA, & NCME Standards (2014)	Implementation Within the Study
Standard 1.1: Establish Intended Uses and Interpretations "The population(s) for which a test is intended should be delimited clearly and the construct or constructs that the test is intended to assess should be described clearly" (p. 23)	The context of the early childhood field was considered during the development of the instrument, including the literature review, expert reviews, and administration of the instrument.
Standard 1.11: Content-Oriented Evidence "When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the constructs the test is intended to measure or the domain it is intended to present." (p. 26)	The content of the instrument was rooted in Evaluation Capacity (EC) theory. Exploratory and confirmatory factor analyses were used for empirical investigation of the EC construct.
Standard 1.13: Evidence Regarding Internal Structure "If the rationale for a test score interpretation for a given use depends on premises about the relationships among test items or among parts of the test, evidence concerning the internal structure of the test should be provided." (p. 27)	Exploratory and confirmatory factor analyses provided evidence for the multidimensionality of the EC construct.
Standard 1.16: Evidence Regarding Relationships With Conceptually Related Constructs "Evidence concerning the constructs represented by other variables, as well as their technical properties, should be presented or cited." (p. 27)	Structural equation modeling provided evidence for the relationships among the individual and organizational factors. Path analysis of multiple groups provided evidence about the influence of the respondents' organizational roles on the relationships among individual and organizational EC factors.
Standard 2.3: Evaluating Reliability/Precision "For each total score, sub-score, or combination of scores that is to be interpreted, estimates of relevant indices of relatability/precision should be reported." (p. 43)	The coefficient alpha was calculated for each factor.

Ethical Considerations

This study received approval from the University of Alberta Research Ethics Board (Reference #: Pro00048028). The study accordingly adheres to the principles of research using human subjects, which can be summarized as respect for persons/autonomy, justice, non-maleficence (do no harm), and beneficence (AERA, 2011; Canadian Institutes of Health Research [CIHR], Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada, 2014). Respect for persons is reflected in the researcher's obligation to protect the rights, welfare, and dignity of the research participants (CIHR et al., 2014). This study provided participants with the freedom to choose whether to participate in the research after being informed of the purpose of the study, by seeking written consent from participants in the expert consultations and expert reviews and from respondents to the EC instrument. Justice, which is an equitable treatment of participants and distribution of benefits and harm (CIHR et al., 2014; Punch, 2013), was met by inviting participants who fit the scope of the study without any unfair or discriminatory exclusion. Beneficence and non-maleficence mean maximizing benefits and minimizing harm for research participants (CIHR et al., 2014; Punch, 2013). In this study, the benefits to participants are intangible, as they are not yet known. Benefits could include enhancing their awareness of the EC construct and the knowledge and skills they may need to conduct evaluations. Finally, potential harm that might come from participation was minimized by informing respondents about the goals of the study and by providing them with a choice regarding participation and the opportunity to stop at any point in the survey with no negative consequences.

Framework and Instrument Development

The EC framework and instrument were developed in three stages and summarized in Figure 2. The process began with operationalizing EC to identify key components and their respective items. The next stage in the process was to draft an initial EC framework and instrument and then conduct two reviews to refine the instrument. These first two stages contributed to gathering validity evidence related to content and consequences. The last stage was to finalize the EC framework and the instrument, which was converted to an electronic format. During each stage, actions were taken to improve the relevance of the EC framework and instrument to the context of the early childhood field adhering to Standard 1.1.

Operationalize EC components	Conduct expert reviews	Revise EC instrument and framework
• Identified components and example items • Conducted two consultations with experts to improve relevance of EC conceptualization to the early childhood field • Created the initial EC instrument and framework	• Conducted a group expert review initial EC instrument • Revised intial instrument, • Conducted individual expert review of the revised EC instrument	• Revised EC instrument, resulting in final version • Converted instrument to electronic format • Revised the EC framework

Figure 2. Steps involved in creating a contextually relevant evaluation capacity instrument.

Operationalization of evaluation capacity components. In the first step, key components and their definitions that are commonly reported in the EC literature were identified (see Chapter 2). It was particularly important to consult experts during this process because the literature review indicated diverse ways of conceptualizing EC, a lack of consensus on the EC

components and their definitions, and a lack of EC frameworks or instruments specifically created for the context of this study. Two committees—six members of the project management team and nine members of the core research team—shared their expertise on measurement, evaluation, and the context of the early childhood field. As a member of the project management team, I facilitated each session. The first consultation, with the project management team, served as a brainstorming session to prepare for the second consultation with the core research team. Together, these consultations were a necessary step toward creating an EC framework and instrument that are relevant to the context of the early childhood field.

> ***Consultation session with the project management team.*** The goal of the consultation with the project management team was to review and evaluate the clarity and appropriateness of the identified EC components and their definitions, with a focus on their relevance to this study's context. The session was divided into three parts: an introduction to the session and explanation of the procedure—an activity to brainstorm about the relevant EC components and their definitions, and a final summary of the results and remaining questions. In the first part I explained the purpose of the session, summarized takeaways from the literature, and identified the gaps that remained. Then I divided the group into three pairs for reviewing and discussing a set of prominent EC components in the literature, such as Organizational Leadership, Individual Attitude, and Individual Motivation. This was followed by a review of the definitions of the components found in the literature. To facilitate this process, I summarized the components and definitions identified in the key frameworks conceptualizing EC or EC building. I also printed the components and their definitions on different coloured papers to allow each team to move them around as they categorized the components and their definitions. This was useful because the same component would often be defined differently by different scholars. Finally, in pairs

and then in the larger group, we discussed questions relating to the strengths and limitations of each component, the clarity of its definition, and its relevance to this study's context.

This session highlighted the importance of context for conceptualizing EC. Overall, the existing conceptualizations of EC included components that appeared less relevant to the context of early childhood, and inconsistencies in the operationalization of the components were identified. For example, motivation in Preskill and Boyle (2008a) concerns the underlying reasons for engaging in building EC, while in Taylor-Ritzler et al. (2013) it concerns the individual's interest in conducting and using evaluations. As discussed in the literature review, the EC components focused on aspects related to individuals (e.g., Individual Skills) or organizations (e.g., Organizational Culture), or both. Some components focused on the capacity to conduct evaluations, use their results, or both. Moreover, some components were considered key to the EC process (e.g., Organizational Resources), while others were considered outcomes of EC (e.g., Individual Readiness). As a team, we agreed that context was not itself a component of EC but instead influenced the conceptualization and measurement of EC components. The components that were finally chosen aligned most closely with the conceptualizations in Cousins and colleagues (2008) and Cousins and colleagues (2014). Features that the group found compelling included highlighting the importance of individuals and organizations and inclusion of components that focus on both the capacity to conduct and the capacity to use evaluations. The session culminated with the group selecting a set of potential EC components that might be relevant to the early childhood context; these would be further discussed with the core research team. The group also generated questions to help guide the consultation with the core research team. Based on our discussion, I created a table summarizing the components that emerged as potentially relevant and a list of questions for the next session.

Consultation session with the core research team. The goal of consulting the core research team was to identify the relevant EC components that best fit the context of the early childhood field, which contributed evidence related to content and consequences. The session was divided into four parts. I began by describing the purpose and procedure for the session (see protocol in Appendix A) and summarizing the key takeaways and gaps in the literature. My initial plan was to begin with individual reviews followed by a group discussion of major points of agreement and disagreement. For the individual reviews, each member would be asked to review and answer questions in a table summarizing the EC components, the respective definitions and example items, and a 3-point scale (1: *Agree*, 2: *Neutral*, 3: *Disagree*) to assess the relevance of each component (see Appendix A). However, after each member briefly reviewed the document, the conversation deviated from my planned agenda. The team members expressed an interest in reconsidering the possible components for inclusion in the instrument. To facilitate this request, I shared the tables and examples of EC frameworks I had identified based on the literature that had been prepared for the previous session with the project management team.

After briefly reviewing the documents and discussing their observations, several members of the team shared their belief that the components identified by Cousins and colleagues were more relevant for the context of the Canadian Federal Government and less so for the early childhood field. For example, while the team agreed that the component Organizational Structure was relevant to EC, they disagreed about its importance to the context of this study, for several reasons. Organizations in the early childhood field have diverse structures, and it would be difficult to measure this component by surveying individuals who may not be familiar with the structure of their organization, especially if they work in a large

organization. One of the main conclusions that emerged from this discussion was that to be relevant to the EC context, an EC framework must fulfill two roles. First, it should identify components at the individual and organizational levels. Second, it should focus on the EC that facilitates conducting and using evaluations. Consistent with this reasoning, the team preferred Preskill's (2008) conceptualization for its clear articulation of individual and organizational EC components, while they preferred Cousins and colleagues' approach for its specification of the relationships among components.

The resulting initial framework and instrument. The literature described in Chapter 2 informed the identification of the six individual and four organizational EC components and their relationships comprising the initial theoretical framework visually represented in Figure 3. The linear aspect of the relationships and the directionality of influence described by the arrows emerged from the theory of planned behaviour to EC. According to the hypothesized framework, achieving organizational mainstreaming, the goal of EC, requires the interaction of both individuals and organizations. To achieve this goal, the organizational components related to Environment, Leadership, and Resources, with each one influencing the next, eventually impact Individual Awareness. The individual components include Awareness, Attitude, Knowledge and Skills, Cultural Competence, Motivation, and Readiness, with each one influencing the next and culminating in Organizational Mainstreaming of Evaluation.
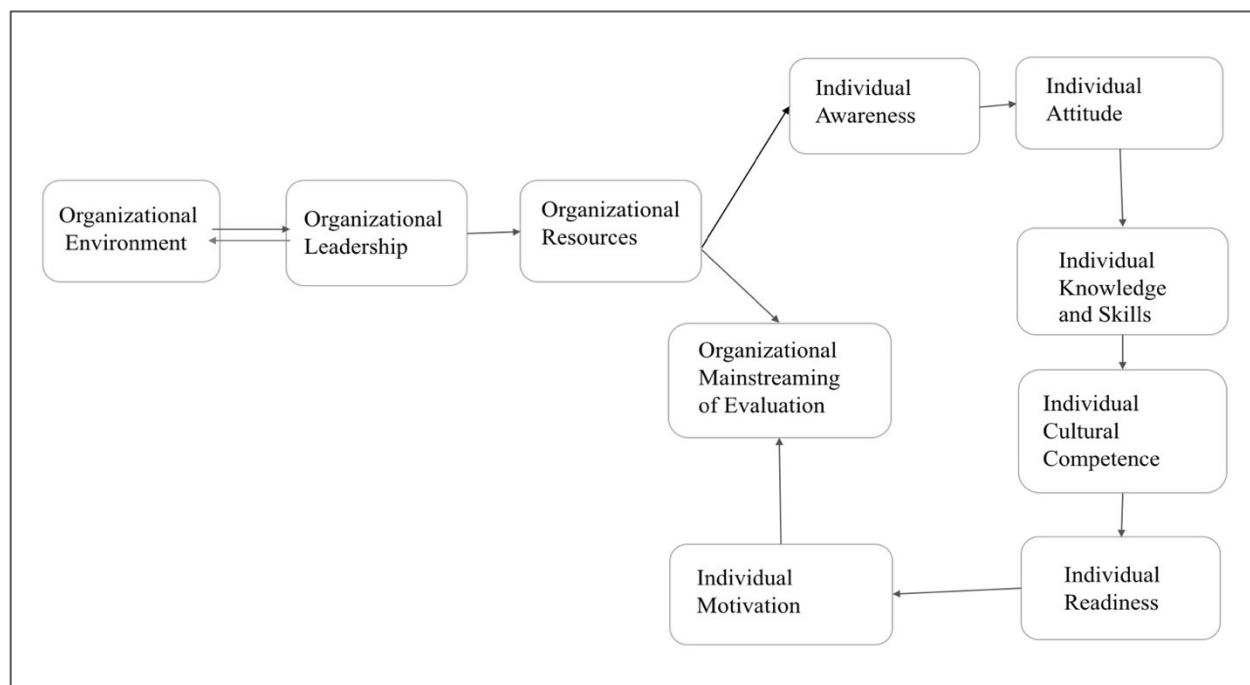
Figure 3. Theorized evaluation capacity framework specific to the early childhood field.

The EC framework reflects the hypothesis that the relationships between and among the EC components, if assessed using cross-sectional data, might demonstrate a series of linear influences. While it may be difficult to predict the nature of all potential relationships, an empirical investigation of these relationships requires balancing comprehensiveness with simplicity (see Bunge, 2012). From an empirical standpoint, to advance EC, it is beneficial to specify relationships that can be assessed based on data, which may allow other scholars to expand this study's findings to capture other potential relationships.

After specifying the components of the initial EC framework, an iterative approach was used to create the initial draft of the EC instrument. Table 3 summarizes the EC components, the number of respective items and examples of items included in the initial draft of the EC instrument. To construct the items, I revisited EC studies identified in the literature review that

operationalized at least one EC component. I extracted items and relevant explanations from existing instruments, checklists, and frameworks that focused on EC or EC building. I also examined relevant instruments measuring one or more of the EC components in other fields, such as leadership and social psychology. After this step, I revised or created items to operationalize each EC component. The first version of the instrument included three sections: Individual EC, Organizational EC, and Respondent's Profile. Each section began with instructions for completing the items. During the process of developing the instrument, I regularly consulted with the project management team to discuss the relevance and clarity of the items to ensure that the EC instrument would meet the needs of the overall project. This iterative process resulted in the initial instrument, which I used for the expert review.

Table 3

Evaluation Capacity Components and Example Corresponding Items for Initial Draft Instrument

Components	Number of Items	Example Items
Individual Evaluation Capacity		
Awareness	1	I have an understanding of evaluation (e.g., key evaluation terms).
Attitude	5	I think evaluation findings can help me better perform my day-to-day work
Motivation	7	I think internal allocation of funding for an evaluation is justified
Knowledge and Skills	4	I know what skill set to look for in an external evaluator
Readiness	8	I see change as an opportunity regardless of whether or not it brings success
Cultural Competency	7	I encourage staff members to openly express their concerns about how we best provide services to different cultural groups.
Organizational Evaluation Capacity		
Organizational Environment	7	Our organization has a practice to involve staff members in making long-term plans.
Leadership	13	The leader/manager of our organization encourages critical thinking and reflection during discussions.
Resources	10	Our organization has the financial support through specific funders to conduct evaluation.
Mainstreaming	10	Has adopted evaluation activities to meet the sponsor's/funder's expectations (adopted from Taylor-Ritzler et al., 2013)

Expert reviews. I used two types of consultations to ascertain the quality and relevance of the instrument for the context of organizations working in the early childhood field. The first type was an expert review in a group format to evaluate the initial draft of the EC instrument in

terms of item clarity, completeness of the components, and appropriateness to the early

childhood context (Gokiert, Noble, & Baugh Littlejohns, 2013). The second type of consultation

had each expert review the instrument after it was revised in light of the group evaluation, and

provide feedback.

   Group review. The EC instrument was reviewed by a group of eight experts who were

purposefully chosen using maximum variation sampling, a sampling process that begins by

establishing criteria to differentiate participants and then selects those who are different with

respect to the specified criteria to increase their diversity (Creswell, 2012). This sampling

method was appropriate because the purpose of the review was to explore the limitations of the

instrument and modifications that might be made, not to achieve consensus among the reviewers.

Consistent with this approach, experts were identified that belonged to three existing project

committees (with a total of 20 committee members), based on their self-reported expertise and

organizational representativeness. The targeted expertise included expertise in evaluations and

EC building, measurement, and the early childhood context. These experts were also

organizationally diverse (e.g., representative of community agencies, government and

nongovernment funding agencies, and academic institutions). Potential participants were

contacted by the overall project's lead researcher; each participant was sent an information letter

and consent form (see Appendix B) inviting them to participate in a 2-hour instrument review

session. The letter contained information about the study, the review process, and their

participant rights, to be signed if they agreed to participate.

   The purpose of the group review was to have experts evaluate the instrument instructions,

the constructs being measured, and the items with respect to clarity, completeness, relevance and

purpose, response format, and expected consequences of the instrument's use. At the beginning

of the review session, the experts had several minutes to review the EC instrument. Then they

were divided into two groups to review and discuss each aspect of the instrument (instructions,

items, scales, and formatting), answering the questions listed in a protocol (Appendix C). After

finishing this 45-min segment, the two groups convened to discuss their conclusions,

agreements, and differences.

The expert review session was captured in several ways. The instrument was divided into

sections and printed in large font on large posters so that each group could write their comments

and suggestions directly on the instrument. I was one of two facilitators who guided the

discussions and asked follow-up questions, as needed, about the clarity of the items, the

completeness of the domains, and the potential consequences. Each group also had one note

taker who focused on capturing the discussion and added relevant contextual descriptions.

During the session, the experts engaged in a rich discussion about the constructs and their

relevance. In addition, they provided specific feedback related to the clarity of the constructs, the

item wording, and the item relevance. Some of the general comments that guided the revisions

involved choosing the constructs that are directly part of EC, making the language of the

instrument relevant to users, shortening the instrument to include only key items, paying

attention to item clarity in terms of measurement (e.g., avoiding double-barreled items—items

that ask more than one question). I used the expert feedback captured in the posters and notes to

revise the instrument. Table 4 summarizes examples of feedback regarding the items and the

resulting changes from the feedback.

The feedback from the expert review led to changes in construct clarity that included the

deletion of awareness as a component. Specifically, the experts agreed that Awareness was less

relevant in the context of this study because organizations in the early childhood field are often

required to evaluate their programs (White & Prentice, 2016). Instead, Individual Attitude was found to be more relevant because it might influence their participation in conducting an evaluation and using it. The experts also agreed that Mainstreaming was a multidimensional construct (Cousins et al., 2014; Preskill & Torres, 2001), which makes it less relevant for an instrument that aims to assess EC at a single point in time (not longitudinally). Together, the expert review process contributed to clearer scale and items and enhancing the contextual relevance of the instrument.

Table 4

Examples of Expert Review Feedback

Type of Feedback	Example of Feedback	Resulting Change
Construct clarity	"Is awareness similar to beliefs? And what is the difference between attitude and beliefs?"	Awareness was deleted and the item was focused on Individual Attitudes regarding conducting and using evaluations.
	"Mainstreaming is difficult to measure especially in the short term and it requires multiple elements"	Mainstreaming was deleted because it is an indirect outcome of building EC over time to change the focus to more immediate outcomes of the motivation to conduct and use evaluations
	"Items related to knowledge and skills should be consistent with the Canadian evaluation competencies"	A section was added that focused on evaluation skills recommended by the Canadian Evaluation Society Standards (Yarbrough et al., 2011).
Scale appropriateness	"not descriptive?"	The scale was revised replacing *not at all*, *to a small extent*, *to some extent*, *to a large extent*, and *NA* with *strongly disagree*, *disagree*, *agree*, *strongly agree*, and *NA*.
		The items were revised to align with the scale.
Item wording	"Split items that ask about two separate ideas"	Items were revised so that each measured only one aspect of the construct.
Item relevance	"Delete 'I think evaluation can be useful in determining how well we are doing' because it is captured by other items"	Items were deleted for which all experts agreed did not add much value.
	"Items asking about respondents' knowledge and skills should take into consideration respondent's role"	Items were revised so that they measure skills based on the respondent's current role in the organization. I also asked about the respondent's role in the demographic section.

Individual reviews. The revised EC instrument was emailed to each expert for another review, asking them to re-evaluate the instrument in terms of item clarity, completeness of the domains, and appropriateness for the early childhood context and to note anything that may not have captured the group review. The experts reviewed the EC instrument using MS Words' "Track Changes" feature. Their comments focused on the relevance of items, their clarity, and suggesting better wording based on their expertise. I revised the instrument to integrate their suggestions, balancing the need to adhere to good measurement practice and relevance to the study's context. For example, in most cases limiting the number of items to three per indicator to make the instrument less time consuming, which may improve respondents' experience as it requires less time to answer. These revisions resulted in the final version of the EC instrument (see Appendix D). Table 5 summarizes the key differences between the three versions of the instrument in terms of the components, the rating scale, and the number of items. The substantial revisions occurred from the initial to revised instrument, which was refined by clarifying items and instructions. The number of components included for measuring individual and organizational EC was reduced from ten in the initial instrument to six in the revised instrument. The change from the first to the second version of the instrument was considerable, as it involved revising the scale and the components that were included as well as enhancing the items' clarity. Revisions going from the second to the third version focused more on the clarity of the instructions and the items.

Table 5

Key Differences Between the Three Versions of the Evaluation Capacity Instrument

Areas	Initial Instrument	Revised Instrument	Final Instrument
Components measured	*6 Individual components*: Awareness, attitude, motivation, knowledge and skills, readiness, and cultural competence	*3 Individual components*: Attitude, motivation, skills	(no change)
	4 Organizational components: Environment, leadership, resources, and mainstreaming	*3 Organizational components*: Culture, leadership, and commitment to evaluation	(no change)
Scale	4-point scale; 1: *Not at all*, 2: *To a small extent*, 3: *To some extent*, 4: *To a great extent*, 88: *Not applicable*	4-point scale; 1: *Strongly disagree*, 2: *Disagree*, 3: *Agree*, 4: *Strongly agree*, 88: *Not applicable*	(no change)
Number of items per section	Individual EC: 35 items	Individual EC: 27 items + 14 items focused on specific skills	Individual EC: 26 items + 16 items focused on specific skills
	Organizational EC: 40 items	Organizational EC: 33 items	(no change)
	Your profile: 12 items	Demographic profile: 16 items	(no change)

The final instrument and revised framework. The final EC instrument has three sections. The first section consists of items assessing self-reports regarding the individual's EC, including the knowledge, skills, and motivations associated with evaluation. This section includes 42 items, 26 of which measure individual EC constructs and 16 measuring specific evaluation skills and their importance for the respondent. The survey instrument requires respondents to rate the 26 items for the individual EC constructs on a four-point Likert scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*). The remaining 16 items assess the level of the respondent's perceived evaluation skills and their importance in the respondent's current job using three-point scales, ranging from 1 (*not at all skilled*) to 3 (*very skilled*) and 1 (*not at all important*) to 3 (*very important*), respectively. These 16 items were not included in the analysis described below because of their interdependence, which violates the independence assumption required for exploratory factor analyses (EFAs), confirmatory factor analyses (CFAs), and structural equation modeling (SEM; Brown, 2006; Kline, 2011).

The second section of the instrument consists of 33 items focused on measuring organizational EC, including respondents' perceptions of their organization's attitude toward evaluation, of the resources invested in evaluation, and of leadership approaches that might enhance or deter EC building efforts. The organizational EC scale uses the same four-point Likert scale. Finally, a third section of the instrument includes demographic information such as gender, age, education level, scope of evaluation experience, job title, job location, and organization type and size.

I next revised the initial EC framework in Figure 3 to reflect the final instrument. The revised framework, presented in Figure 4, includes six instead of ten components. Three individual components (Awareness, Cultural Competence, Readiness) and only one

organizational component (Mainstreaming of Evaluation) were deleted. Finally, the relationship between Organizational Culture and Organizational Leadership has changed from a dyadic relationship to one in which Organizational Culture is theorized to influence Organizational Leadership this study's across sectional design does not make it possible to assess the dyadic aspect. The framework displays hypothetical linear relationships between and among the EC components, which were statistically tested using SEM. The specific relationships and their potential direction were identified based on understandings generated from the literature (see Chapter 2) and are further discussed in Chapter 5.
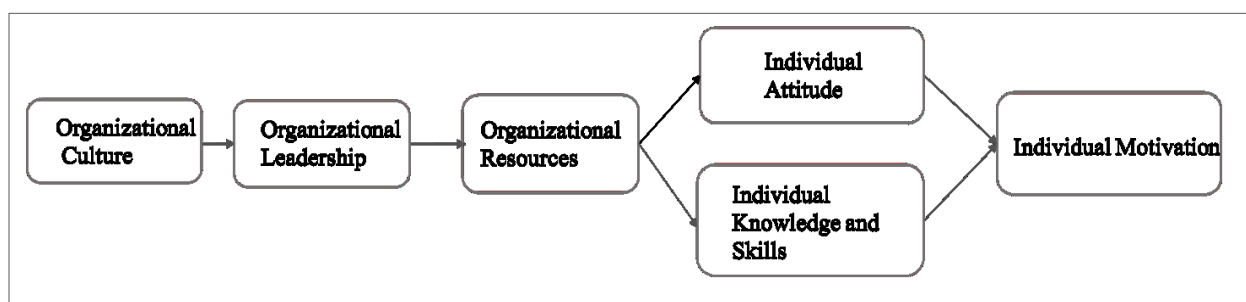
Figure 4. Revised evaluation capacity framework relevant to the early childhood field.

Data Collection

　　Sampling strategy. A nonprobability sampling strategy was used—specifically, purposive sampling, which is appropriate when seeking members of predefined groups (Kemper, Stringfield, & Teddlie, 2003). To collect quantitative data, participants working in the early childhood field (e.g., in community organizations, funding agencies, government agencies, academic departments) in Alberta, Canada were recruited. Specifically, the targeted population included individuals who were responsible for conducting evaluations, supervising evaluations performed by external evaluators, or using evaluation results for decision-making.

The sample for this study consisted of 332 respondents working in the early childhood field. Data were collected in two stages. In the first stage (January to February 2016), a sample of 101 early childhood stakeholders responded. The second stage (May to June 2016) involved 228 respondents. The respondents in the first sample were specifically targeted because of their roles as decision makers in their organizations, which would allow one of the goals of the overall project to be addressed. Because of this focus, the first sample exhibited less diversity than the second sample in terms of professional roles. Taken together, however, the two samples provided a better representation of the individuals and organizations in the early childhood field in Alberta, which led to the decision to merge the samples for the analysis.

The valid sample used for analysis included 329 respondents after removal of three respondents that skipped 80% of the items. Because the study sample was drawn from an unknown population size, I was not able to calculate a response rate. There is no clear consensus on the required minimum sample size when conducting SEM (Schumacker & Lomax, 2012); however, a general recommended sample size is 200 (Kline, 2011). The ideal sample size depends on several considerations, such as the number of latent factors, the number of observed items, and the intended use and potential consequences of the resulting conclusions (AERA et al., 2014; Kline, 2011). For the present study, a sample size of 329 was large enough to draw robust conclusions for the initial validation of the EC instrument.

Survey administration. The EC instrument was administered using an online platform, FluidSurveys (FluidSurveys, 2015). Web-based surveys have several benefits compared to traditional, paper-based surveys, including a low cost of administration, easy access via an emailed link, and greater accuracy because there is no need for data entry (Dillman, Smyth, &

Christian, 2014). Although access to technology is often a challenge associated with web-based administration (Dillman et al., 2014), this was not an issue for the present study.

In the first stage, the lead researcher for the overall project sent email invitations to possible participants (see Appendix B). There were two types of emails inviting individuals to complete the EC instrument and to attend events called *Evaluation Capacity Building Forums*, which were held to discuss evaluation capacity needs within the early childhood field (for a detailed description of the forums, see Gokiert et al., 2017). The early childhood stakeholders who were initially invited to complete the survey and attend the community dialogues were considered leaders in their field. Two email reminders to complete the EC instrument were sent during the second and fourth weeks after the initial invitation. For the second stage, three months later, the sampling strategy was broadened to allow a better understanding of EC in the early childhood field. The demographic sections of the instrument were modified to address the needs of the overall project. The lead researcher sent another invitation email to early childhood community organizations in Alberta, with a request that the instrument be completed by their staff. Both emails described the purpose of the study and the EC instrument, and then it directed recipients to follow the provided link to the online instrument. Respondents were advised about the procedure of providing informed consent (see Appendix B) before starting the instrument and were asked to provide their consent by submitting a response. Two email reminders to complete the instrument were sent during the second and fourth weeks after the initial email.

For the second administration, a screening question asking respondents about their participation in evaluation activities was added to the EC instrument. Participation in evaluation activities was defined broadly, to include gathering, analyzing, summarizing, interpreting, and reporting evaluative data/information. The purpose of the screening question was to make the EC

survey relevant to the respondents (by asking them to skip certain items) and less time-consuming. However, this change led to systematically missing responses for several items, which required deciding whether to include these respondents in the sample during the analysis. These responses were finally included, given the size of the sample; the rational for this decision is further discussed in the analysis section of this chapter.

Participant characteristics. The characteristics of the respondents, organized by the overall sample and the two samples for the two stages, are reported in Table 6. The majority of the respondents were female (92%), and the majority of the respondents' ages ranged between 40 and 59 years (58%). The respondents' professional roles were somewhat diverse in the overall sample, including 33.8% in management positions, 15% in leadership roles, and 22.5% in front-line positions. The majority (87%) of the respondents had some evaluation experience, with 44.8% reporting five or more years of evaluation experience, and more than half (59.2%) of the respondents reported at least an undergraduate degree as the highest level of education they had completed.

Table 6.

Study Participants' Demographics

Item	Overall sample (*N* = 329)		Sample 1 (*N* = 101)		Sample 2 (*N* = 228)	
	N	Valid %	*N*	Valid %	*N*	Valid %
Gender						
Female	301	92.3	90	89.1	211	93.8
Male	17	5.2	10	9.8	7	3.1
Professional role						
Leadership (e.g., CEO, executive director)	48	15.0	27	27.6	21	9.5
Management (e.g., program director, supervisor)	108	33.8	26	26.5	82	36.9
Front-line (e.g., works directly with clients)	72	22.5	1	1.0	71	32.0
Support staff	17	5.3	5	5.1	12	5.4
Researcher/academic	7	2.2	6	6.1	1	.5
Educator	30	9.4	10	10.2	20	9.0
Evaluator (e.g., evaluation consultant, internal evaluator)	22	6.9	14	14.3	8	3.6
Other	16	5.0	9	9.2	7	3.2
Years of evaluation experience						
None	43	13.2	7	6.9	36	16.0
Less than 1 year	39	12.0	7	6.9	32	14.2
1 to 4 years	98	30.1	28	27.7	70	31.1
5 years or more	146	44.8	59	58.4	87	38.7
Highest level of education completed						
High school diploma or equivalent	12	3.7	1	1.0	11	4.9
Post-secondary certificate or diploma	121	37.1	17	16.8	104	46.2
Undergraduate degree	87	26.7	30	29.7	57	25.3
Graduate degree	106	32.5	53	52.5	53	23.6
Age (in years)						
29 years or less	29	9.2	4	4.0	25	11.6
30–39 years	75	23.9	16	16.2	59	27.4
40–49 years	91	29.0	31	31.3	60	27.9
50–59 years	91	29.0	37	37.4	54	25.1
60 years or more	28	8.9	11	11.1	17	7.9

Data Analysis

The data analysis was conducted in four steps. First, the data were prepared and preliminary analyses were conducted. This was followed by EFA, CFA, SEM, and multi-group path analysis to answer the research questions. The answers to the second and third questions built on the results of the previous question. SPSS software version 23 (IBM Corp., 2015) was used for the preliminary analyses, and Mplus 7.4 (Muthén & Muthén, 2015) for the EFA, CFA, and SEM analyses.

Data preparation and preliminary analyses. Prior to the analyses, the data were checked for missing values and inconsistencies (Fabrigar, 2012). The data had two types of missing values: values that were missing for unknown reasons and values that were systematically missing because of an automatic skip based on the respondent's response to the screening. Three participants (0.01% of the sample) missed 80% or more of all items; their responses were therefore removed from the subsequent analyses. Fifty-two respondents (15%) had systematic missing data because the respondents indicated, in response to the screening question, that they did not contribute to evaluation activities in their present position. The items affected by the screening question included nine items in the individual EC scale and 12 items in the organizational EC scale. Given the already small sample size for adequately conducting an SEM analysis, and after the descriptive statistics for this group were checked, all of these respondents were included in the analysis. This decision had implications for the estimator that was chosen for the EFA, CFA, and SEM—the full-information maximum likelihood with robust standard error (MLR) estimator—as it takes missing values into account.

The data were also investigated to verify univariate normality, and new indicators were created as needed (Tabachnick & Fidell, 2006). I analyzed the items of the individual and organizational EC scales separately, consistent with the theory that contributed to the development of the instrument. As expected for survey data, the descriptive statistics for the individual and organizational EC items (Appendix E) indicated that the data were skewed toward positive responses. Since the data revealed a violation of univariate normality, the MLR estimator was also chosen to address this violation (Li, 2016).

Factor analysis. EFA and CFA were used to answer the first research question, which focused on the underlying constructs and relationships comprising EC in the early childhood field. The purpose of EFA is to ascertain the nature and number of latent factors accounting for the variation and covariation that exist among items (Thompson, 2004). EFA allows the variance of each measured item to be partitioned into a common variance and a unique variance, with the goal of achieving a scale that has a simple structure (Floyd & Widaman, 1995). Determining the nature and number of factors provided a tool that can be used to eliminate items that do not adequately contribute to measuring the latent factors. EFA was appropriate because empirical understanding of the EC conceptualization is not firm. For this question, the theoretical a priori discussed in the literature review chapter was set aside to inductively determine the nature of the factors using the data (Thompson, 2004; Treiblmaier & Filzmoser, 2010).

The 26 individual and 33 organizational EC items were analyzed separately with EFA, using the MLR estimator to adjust for item non-normality and missing values and a Geomin rotation to adjust for inter-factor correlations (Treiblmaier & Filzmoser, 2010). The number of factors was determined by comparing the goodness of fit, the comparative indices of six potential

models, and the Cattell-Nelson-Gorsuch (Gorsuch, 1983) scree plot, along with the theoretical interpretation of the factors.

For goodness-of-fit indices, the root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), standardized root mean square residual (SRMR), and chi-square test of model fit (Brown, 2006; Kline, 2011) were used. The RMSEA is a parsimony-corrected index that reflects closer fit when its value is close to 0.07 (Steiger, 2007). An RMSEA less than 0.08 represents moderate fit, while an RMSEA less than 0.05 represents good fit (Brown, 2006; Kline, 2011; Steiger, 2007). The CFI measures a model's relative improvement in fit compared to the baseline model, assuming uncorrelated indicators; it ranges from 0 to 1.0, with a value greater than 0.9 representing an acceptable fit and a value greater than 0.95 representing a good fit (Hu & Bentler, 1998; Kline, 2011). The TLI is similar to the CFI, but less affected by sample size (Hu & Bentler, 1998; Kline, 2011). The SRMR measures the mean absolute residual correlation, with values close to 0 being best and values smaller than 0.08 representing a good fit (Hu & Bentler, 1998; Kline, 2011).

The comparative indices of Akaike information criterion (AIC; Akaike, 1987) and Schwarz's Bayesian information criterion (BIC) are statistics of goodness-of-fit measures that allow comparing EFA models with different numbers of factors while considering the complexity of the models (Brown, 2006). AIC and BIC are similar, and their interpretation is derived from comparing their values as the model changes; the best model is the one with the smallest indices (Field, 2013).

During the EFA, items with loadings of .30 or less ($\leq$.30) and those that loaded equally on multiple factors were eliminated. Factors that were measured by only one or two items were also eliminated, as the items lack stability for measuring the factor (Floyd & Widaman, 1995).

The decisions to eliminate items and factors were made in accordance with the statistical results and the theoretical understanding of EC.

CFA, which is a multivariate statistical technique, tests the factor structure an instrument purports to measure (Brown, 2006; Kline, 2013). CFA was used to examine the relationships between the items and the factors to investigate whether the scales measure distinct constructs consistent with the EC framework. The CFA establishing a good-fitting measurement model preceded the SEM, because without first conducting CFA, it is difficult to ascertain the reason for poor-fitting structural solutions (Brown, 2006). A poor-fitting SEM model can be due to the measurement solution defining the relationships between variables and constructs and/or the structural solution defining the relationships among the factors (Brown, 2006).

Similar to the EFA, the individual and organizational EC scales were analyzed separately in the CFA, using an MLR estimator. This estimator provides a robust approach to correcting for distortions in fit indices and standard errors caused by multivariate non-normality (Enders, 2001). The CFA model was evaluated using the same fit indices as were used for the EFA.

Reliability analysis was used to examine the internal consistency of the items measuring individual and organizational factors. Reliability was determined by calculating the coefficient alpha (Cronbach, 1951), which is widely used to assess internal consistency (DeVellis, 2012). A coefficient alpha between .8 and .9 indicates high internal consistency, and a coefficient alpha between .7 and .8 indicates moderate internal consistency.

Structural equation modeling. SEM was used to answer the second research question, concerning the extent to which the data support the theorized EC within the early childhood field. This analysis was used to investigate the "hypothesized pattern of directional and non-directional linear relationships" (MacCallum & Austin, 2000, p. 202) among the individual and

organizational EC factors identified by CFA. The model data fit was evaluated using the chi-square test of model fit (Brown, 2006; Kline, 2011), and the goodness of fit was evaluated using the same fit indices as in the EFA and CFA analyses, which included CFI, RMSEA, and SRMR. Indirect paths were tested with a bias-corrected bootstrap confidence interval (CI; Preacher & Hayes, 2008).

An alternative SEM model that differed from the main SEM model only in the direction of influence between organizational leadership and organizational culture was tested, which contributed validity evidence regarding the model's internal structure and relationship to other variables. Initially, a model with culture influencing leadership was tested, and in the alternative model, leadership influencing culture was tested. The reason for testing this relationship was that although, as reported in the literature review (Chapter 2), the leadership literature clearly discusses a bidirectional influence between leadership and culture, there is not as much empirical evidence investigating this claim.

Multi-group path analysis. Multi-group path analysis was used to answer the third research question, which focuses on whether the structural model identified in the previous step differs for different professional roles. To examine whether the path coefficients are different across professional roles, a chi-square difference test for multi-group path analysis was conducted (Sauer & Dick, 1993). Multi-group path analysis was used instead of multi-group SEM because the latter requires a larger sample size than was available in this study. For the multi-group path analyses, factor scores saved from the final SEM model were used. Given that the factor scores are assumed to be normally distributed, maximum likelihood rather than MLR was used to estimate the model parameters. Two multi-group path models were estimated: one with path coefficients freed for the two professional roles, and one with equal path coefficients

for respondents who indicated that they had a professional leadership role and those who indicated non-leadership professional roles. The differences between the chi-square values and the degrees of freedom for the two models were used to determine whether the path coefficients differ significantly across the two professional roles.

Chapter 4: Findings

This chapter reports the results of the psychometric analysis of the EC instrument. First, the results of the EFA are reported, followed by an additional analysis of the factors as interim preparation prior to conducting the CFAs for the individual and organizational survey items. Finally, the results of testing the EC model with SEM are reported. Together, these analyses provide evidence that supports validation of the inferences generated from the EC instrument for the early childhood field.

Exploratory Factor Analysis

EFA was used to identify the underlying factor patterns for the individual and organizational EC items. While the instrument was built based on the theoretical understanding in the existing literature, there is no firm apriori theory because the EC field is still young and this is a newly administered instrument (Floyd & Widaman, 1995). The EFA resulted in two distinct models: one for the individual EC scale and another for the organizational EC scale.

Individual evaluation capacity findings. Although a three-factor structure was expected, based on the literature review that informed the development of the EC instrument, the analysis examined six solutions with one, two, three, four, five, and six factors. The process of identifying the factor structure with the best fit involved examining the factor solutions summarized in Table 7 that lacked support based on the fit-indices (Brown, 2006) and the Cattell-Nelson-Gorsuch scree plot (Gorsuch, 1983) presented in Figure 5, the factor loadings, and item content.

Table 7

Goodness of Fit and Comparative Indices for Initial Six Exploratory Factor Analysis Solutions for Individual Evaluation Capacity

Solution	χ^2	df	CFI	TLI	RMSEA	SRMR	AIC	BIC	Adjusted BIC
1-factor	1446.96	299	. 59	. 60	.11	.10	11088.69	11384.78	11137.37
2-factor	1033.08	274	.73	.68	.09	.07	10626.33	11017.33	10690.61
3-factor	794.51	250	.80	.74	.08	.05	10358.15	10840.25	10437.40
4-factor	672.09	227	.84	.77	.08	.04	10211.22	10780.63	10304.83
5-factor	525.75	205	.88	.82	.07	.04	10089.51	10742.43	10196.85
6-factor	452.05	184	.90	.83	.07	.03	10015.09	10747.73	10135.53

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; Adjusted BIC = Bayesian information criterion adjusted for model overfitting.
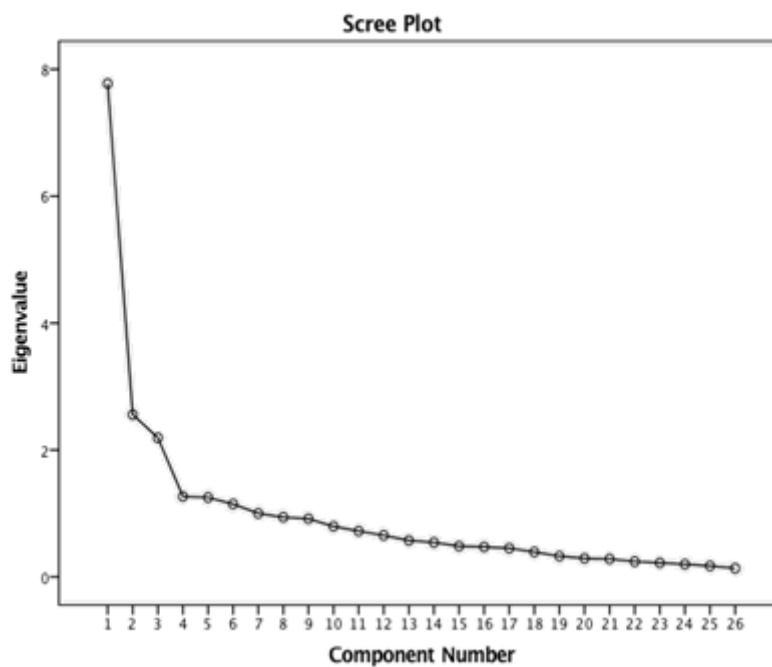
Figure 5. Scree plot the individual evaluation capacity scale.

The fit and comparative indices for one- and two- factor solutions indicated a poor fit because almost all fit indices fell outside of the recommended range (i.e., SRMR <.08, RMSEA <.08, TLI >.90, CFI >.90). The indices provided stronger support for six-factor solutions than for three-, four- and five- factor solutions. An examination of the items measuring five- and six-factor solutions, however, indicated that the fifth and sixth factor were measured by only one or two items, which is not adequate for measuring latent factors (Brown, 2006). Given this limitation, the five- and six- factor solutions were eliminated. The three- and four-factor solutions thus provided better alternatives for further investigation.

Deciding between the three- and four- factor solutions involved inspecting the Cattell-Nelson-Gorsuch scree plot (Gorsuch, 1983). The Examination of the scree plot presented in Figure 5 provided support for the three-factor solution, given the eigenvalues of the first three factors account for most of the variance explained. The content of the items defining the three factors were aligned with theoretical understanding within the literature. Together, this evidence, therefore, supported the conclusion that the three-factor structure provides an overall better fit with the data.

After determining that the three-factor structure fit the data best, the items were evaluated based on their factor loadings and content. To achieve a simple and stable factor structure, items with factor loadings less than .3 were removed because they may not have assessed the factors with strong reliability. Items that loaded equally on two or more factors were first examined in terms of their content to investigate whether they measured multiple dimensions. Items such as "I have a clear sense about the benefits of evaluation" were eliminated because they measured both Individual Skills and Individual Attitude, which mad them less reliable in assessing these two factors. Going through this process led to a three-factor structure measured by 16 items.

These 16 items were re-examined with the three-factor solution using an MLR estimator with a Geomin rotation. The final EFA pattern matrix loading is presented in Table 8. According to the overall goodness-of-fit indices, the final EFA solution of a three-factor individual EC had a good fit with the data, SRMR = .041, RMSEA = .082 (90% CI [.07, .094]), TLI = .82, CFI = .90. The chi-square test of model fit, χ^2 (75, N = 329) = 240.87, p < .001, was significant, indicating a poor fit; however, as this test is stringent and requires a perfect model fit (Brown, 2006), it was not used to evaluate the goodness of fit. Labels for the factors were chosen based on the items representing them. The first factor (items 1 to 6) is Individuals' Attitude toward Evaluation, with factor loadings ranging between .59 and .77. The second factor (items 7 to 12) is Individuals' Motivation to Conduct and Use Evaluations, with factor loadings ranging between .40 and .70. The third factor (items 13 to 16) is Individuals' Evaluation Skills, with factor loadings ranging from .66 to .85. The EFA yielded significant correlations between Individuals' Attitude and Individual's Motivation to Conduct and Use Evaluations, r = .48, p < .05, and between Individuals' Motivation to Conduct and Use Evaluations and Individuals' Evaluation Skills, r = .46, p < .05.

Table 8

Factor Loadings for Exploratory Factor Analysis of the Individual Evaluation Capacity Scale with a Geomin Rotation

Items	Individuals' Attitude Toward Evaluation	Individuals' Motivation to Conduct and Use Evaluations	Individuals' Evaluation Skills
1. think evaluation can be useful in determining the effectiveness of my organization.	**.60***	.15	.01
2. I think evaluation findings can be beneficial to my sponsor(s)/funder(s).	**.70***	.13	-.03

	Factor 1	Factor 2	Factor 3
3. I think evaluation can improve transparency in an organization.	**.77***	-.00	.15
4. I would be concerned if an organization doesn't evaluate its activities.	**.59***	-.03	.16
5. I have an ethical responsibility to participate in evaluation as needed.	**.68***	.09	.01
6. I think it is important that staff members get involved in evaluation.	**.60***	.17	-.03
7. I think internal allocation of funding for an evaluation is justified.	.01	**.70***	.05
8. I think credentialed training (e.g., course, certificate, and diploma) in evaluation is important.	-.08	**.54***	-.02
9. I think non-credentialed training (e.g., webinar, coaching) in evaluation is important.	-.01	**.57***	-.02
10. I think external funding for an evaluation is justified.	.01	**.58***	.03
11. I am open to staff being provided with the opportunities to learn the skills necessary to conduct evaluation.	.02	**.60***	-.01
12. I am open to adopting new ideas in my day-to-day activities based on evaluation findings.	.22	**.40***	.04
13. I know what skills to look for in an external evaluator.	-.04	.10	**.74***
14. I have the skills to oversee an external evaluator.	.02	.01	**.85***
15. I know how to use evaluation findings in decision-making.	.22	0	**.66***
16. I know how to make organizational level changes based on evaluation findings.	.18	-.00	**.71***

Note. Factor loadings ≥ .40 are in boldface. Decision to delete items were based on item content and statistical results.
* *p* > .05.

Organizational evaluation capacity findings. Like the items in the individual EC scale,

the 33 items in the organizational EC scale involved first evaluating six solutions with one, two,

three, four, five, and six factors to identify the factor structure with the best-fit. Similarly, the

process of identifying the factor structure with the best fit involved examining the factor

solutions summarized in Table 9 that lacked support based on the fit-indices (Brown, 2006) and

the Cattell-Nelson-Gorsuch scree plot (Gorsuch, 1983) presented in Figure 6, the factor loadings,

and item content.

Table 9

Goodness of Fit and Comparative Indices for Initial Six Exploratory Factor Analysis Solutions
for Organizational Evaluation Capacity

Solution	χ^2	*df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	Adjusted BIC
1-factor	2487.747	495	.64	.62	.11	.11	15962.87	16337.17	16023.15
2-factor	1655.69	463	.79	.75	.09	.06	14971.82	15467.10	15051.58
3-factor	1284.53	432	.85	.81	.08	.05	14554.61	15167.09	14653.24
4-factor	1140.07	402	.87	.83	.08	.04	14332.89	15058.80	14449.79
5-factor	934.96	373	.90	.86	.07	.04	14167.67	15003.22	14302.23
6-factor	729.85	345	.93	.89	.06	.03	14025.07	14966.48	14176.67

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index;
RMSEA = root mean square error of approximation; SRMR = standardized root mean square
residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; Adjusted
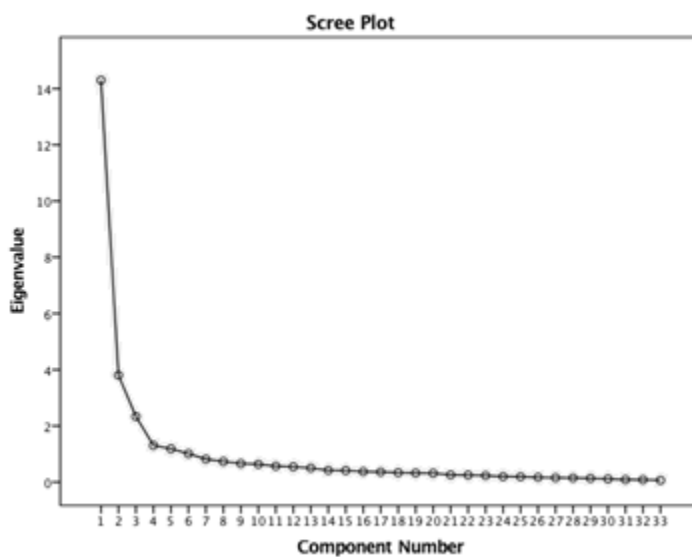BIC = Bayesian information criterion adjusted for model overfitting.

Figure 6. Scree plot of the organizational evaluation capacity scale.

As in the individual EC scale, the fit and comparative indices, reported in Table 9, for one- and two- factor solutions indicated a poor fit because nearly all fit indices fell outside the recommended range (i.e., SRMR <.08, RMSEA <.08, TLI >.90, CFI >.90). The five- and six-factor solutions were eliminated because the fifth and sixth factors had no item with loadings above .30, which indicated lower reliability in measuring those two dimensions. The three- and four- factor solutions were, therefore, further investigated. The Cattell-Nelson-Gorsuch scree plot shown in Figure 6 and the content of the items provided support for the three-factor solutions. Similarly, the item content and examined evidence supported the conclusion that the three-factor structure provided a better fit with the data.

After deciding on the three-factor structure fit the data best, the quality of the items factor loadings was evaluated, and a similar approach to Individual factor structure was used to achieve a simple and stable factor structure. Items with factor loadings less than .30 were eliminated because they may not measure the factor reliably. Then items that loaded equally on two or more factors were examined in terms of content. Similarly, items such as "Leadership has the skills necessary to undertake an evaluation" were eliminated because they measure aspects of Organizational Leadership and Organizational Culture, which may weaken the validity of the conclusions made based on the data. This process led to a three-factor structure measured by 12 items.

These remaining 12 items were re-examined with a three-factor solution using an MLR estimator with an oblique rotation. According to the overall goodness-of-fit indices, the final EFA solution of a three-factor organizational EC had a good fit with the data, SRMR = .022, RMSEA = .038 (90% CI [.009, .060]), TLI = .98, CFI = .98. Table 10 presents the final EFA pattern matrix loading for the three-factor solution. The first organizational factor,

Organizational Culture (items 1 to 4), refers to the extent to which an environment enables evaluative thinking and collaboration, with factor loadings ranging from .59 to .91. The second factor, Organizational Commitment to Evaluation (items 5 to 8), refers to the resources, relationships, and expertise that organizations put in place to facilitate conducting and using evaluations. This factor had loadings that ranged from .57 to .99. The last factor, Organizational Leadership (items 9 to 12), refers to organizational leadership that is open to ideas and encourages risk taking, with factor loadings ranging from .58 to .80. The analysis yielded significant correlations between all three organizational factors: Organizational Culture had a low correlation with Organizational Commitment to Evaluation, $r = .24$, $p < .5$, and a moderate correlation with Organizational Leadership, $r = .68$, $p < .05$, and Organizational Commitment to Evaluation had a low correlation with Organizational Leadership, $r = .32$, $p < .05$.

In sum, the EFA results indicate that three distinct factors underlie the individual EC scale and three distinct factors underlie the organizational EC scale. These results provided a basis for evaluating the measurement models for the individual and organizational EC scales using CFA.

Table 10

Factor Loadings from Exploratory Factor Analysis of the Organizational Evaluation Capacity Scale with an Oblique Rotation

Items	Organizational Culture	Organizational Commitment to Evaluation	Organizational Leadership
My organization …			
1. encourages staff to express their opinions.	**.77***	-.03	.08
2. involves staff when making long-term plans.	**.85***	-.02	-.01
3. gives staff the opportunity to reflect on organizational goals.	**.91***	.05	-.09
4. reviews its mission, vision, and values with staff.	**.59***	.19*	.08
5. dedicates funds to conduct an evaluation.	.04	**.78***	.12
6. dedicates funds to ensure ongoing evaluation.	.00	**.99***	.00
7. has resources in place to undertake evaluation on an ongoing basis.	-.03	**.82***	.04
8. has the commitment from external stakeholders to ensure evaluation sustainability.	.06	**.57***	.14
The leadership of my organization...			
9. builds ideas in collaboration with staff members.	.21*	-.02	**.64***
10. resolves inter-personal conflicts in a positive manner.	-.01	.14	**.80***
11. celebrates staff members' achievements.	-.01	.06	**.78***
12. promotes evaluative thinking.	.04	.28*	**.58***

Note. Factor loadings ≥ .40 are in boldface. Decision to delete items were based on item content and statistical results.

* $p > .05$.

Confirmatory Factor Analysis

CFAs were conducted to evaluate the measurement models based on the EFA solutions for the individual and organizational EC scales. The findings reveal that both measurement models have a good fit with the data, which provided supporting validity evidence related to internal structure.

Individual evaluation capacity scale. A CFA was conducted with the 16 individual EC items using a full-information MLR estimator to examine whether these items measure the scale adequately. Based on modification indices and an examination of item content, the residuals of two items measuring Individuals' Evaluation Skills, "I know what skills to look for in an external evaluator" and "I have the skills to oversee an external evaluator," were allowed to freely correlate with each other. Both items addressed an aspect of hiring and overseeing an external evaluator, and their inter-item correlation was statistically significant ($p < .05$) and moderately high, which could imply the existence of another factor. However, since two items are insufficient to measure a factor, Brown (2006) recommends allowing the items' residuals to freely correlate to account for the relationship.

Consistent with the EFA findings, the overall goodness-of-fit indices generated by the CFA, SRMR = .050, RMSEA = .060 (90% CI [.049, .071]), TLI = .90, CFI =. 92, indicated a good fit with the data for the three-factor individual EC model, which is better than the one-factor models (see Table 11). The chi-square test of model fit, $\chi^2(100, N = 329) = 217.5$, $p < .001$, was significant, indicating a poor fit; however, because the chi-square test is stringent and requires a perfect model fit (Brown, 2006), it was not used to evaluate the EC measurement model's goodness of fit.

Table 11

Goodness of Fit and Comparative Indices Obtained from two Confirmatory Factor Analysis Solutions for Individual Evaluation Capacity

Solution	χ^2	*df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	Adjusted BIC
1-factor	529.65	103	.71	.67	.11	.11	6554.04	6740.04	6584.62
3-factor	217.56	100	.92	.90	.06	.05	14554.61	15167.09	14653.24

Note. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; Adjusted BIC = Bayesian information criterion adjusted for model overfitting.

The results of the CFA testing a three-factor measurement model are presented in Table 12. The first factor, Individuals' Attitude Toward Evaluation, was measured by six items with loadings ranging from .60 to .80, and a coefficient alpha of .85 indicated good internal consistency among these items. The second factor, Individuals' Evaluation Skills, loaded on four items with factor loadings ranging from .62 and .83 and good internal consistency, as reflected in the coefficient alpha of .86. The third factor, Individuals' Motivation to Conduct and Use Evaluations, loaded on six items with factor loadings ranging between .47 and .70 and a moderate internal consistency, according to the coefficient alpha of .73. The CFA yielded significant latent correlations between the three individual EC factors; Individuals' Attitude Toward Evaluation had a low correlation with Individuals' Evaluation Skills, $r = .24$, $p < .05$, and a moderate correlation with Individuals' Motivation to Conduct and Use Evaluations, $r = .68$, $p < .5$; and Individuals' Evaluation Skills had a low correlation with Individuals' Motivation to Conduct and Use Evaluations, $r = .32$, $p < .05$. These inter-factor correlation results provide support for the conceptual separation of the three individual EC factors.

Table 12

Standardized Factor Loadings Obtained from Confirmatory Factor Analysis for the Individual Evaluation Capacity Scale

Items	Individuals' Attitude Toward Evaluation $\alpha = .85$	Individuals' Evaluation Skills $\alpha = .86$	Individuals' Motivation to Conduct and Use Evaluations $\alpha = .73$
1. I think evaluation can be useful in determining the effectiveness of my organization.	.68		
2. I think evaluation findings can be beneficial to my sponsor(s)/funder(s).	.77		
3. I think evaluation can improve transparency in an organization.	.80		
4. I would be concerned if an organization doesn't evaluate its activities.	.60		
5. I have an ethical responsibility to participate in evaluation as needed.	.74		
6. I think it is important that staff members get involved in evaluation.	.69		
7. I know what skills to look for in an external evaluator.		.62	
8. I have the skills to oversee an external evaluator.		.71	
9. I know how to use evaluation findings in decision-making.		.81	
10. I know how to make organizational level changes based on evaluation findings.		.83	
11. I think internal allocation of funding for an evaluation is justified.			.70
12. I think credentialed training (e.g., course, certificate, diploma) in evaluation is important.			.47
13. I think non-credentialed training (e.g., webinar, coaching) in evaluation is important.			.56
14. I think external funding for an evaluation is justified.			.57
15. I am open to staff being provided with the opportunities to learn the skills necessary to conduct evaluation.			.62
16. I am open to adopting new ideas in my day-to-day activities based on evaluation findings.			.58

Note. α = coefficient alpha; * $p > .05$.

Organizational evaluation capacity scale. As was done for the individual EC scale, a CFA was conducted with the 12 organizational items using a full information MLR estimator to test whether the items measure the organizational EC scale adequately. The fit indices, SRMR = .057, RMSEA = .055 (90% CI [.040–.069]), TLI =. 95, CFI = .96, indicated a good model fit for a three-factor structure, better than the one-factor models (see Table 13). The three-factor organizational EC model consisted of 12 items with factor loadings higher than .61, all statistically significant, $p < .01$. Table 14 presents the items, the standardized loadings, and the coefficient alpha.

Table 13

Goodness of Fit and Comparative Indices for two Confirmatory Factor Analysis Solutions for Organizational Evaluation Capacity

Solution	χ^2	df	CFI	TLI	RMSEA	SRMR	AIC	BIC	Adjusted BIC
1-factor	599.94	66	.67	.61	.16	.14	6495.78	6640.03	6519.49
3-factor	117.93	60	.96	.95	.055	.057	14554.61	5808.34	5649.54

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; Adjusted BIC = Bayesian information criterion adjusted for model overfitting.

The first factor, Organizational Culture, loaded on four items with factor loadings ranging between .72 and .85 and a coefficient alpha of .88. The second factor, Organizational Commitment to Evaluation, loaded on four items with factor loadings ranging from .61 to .97 and a coefficient alpha of .88. The last factor, Organizational Leadership, loaded on four items with factor loadings ranging from .70 to .83 and a coefficient alpha of .86. The internal consistency based on the coefficient alpha of the three factors indicated good internal consistency among each factor's items. The analysis yielded significant latent correlations among the three organizational factors; Organizational Leadership had a low correlation with

Organizational Commitment to Evaluation, $r = .49$, $p < .05$, and a high correlation with

Organizational Culture, $r = .81$, $p < .05$, and Organizational Commitment to Evaluation had a

moderate correlation with Organizational Culture, $r = .46$, $p < .05$. The inter-factor correlations

provide evidence warranting the conceptual separation between the three organizational EC

factors.

In sum, the CFA resulted in two measurement models that adequately measure the

individual and organizational EC scales. These findings are consistent with the underlying

theoretical understanding of EC, which provides additional validity evidence related to content

and internal structure. Next, the relationships between the individual and organizational EC

factors resulting from these analyses were used to test the theorized EC model with an SEM

analysis.

Table 14

Standardized Factor Loadings From Confirmatory Factor Analysis of the Organizational Evaluation Capacity Scale

Items	Organizational Culture α=.88	Organizational Commitment to Evaluation α=.88	Organizational Leadership α=.86
My organization …			
1. encourages staff to express their opinions.	.81		
2. involves staff when making long-term plans.	.83		
3. gives staff the opportunity to reflect on organizational goals.	.85		
4. reviews its mission, vision, and values with staff.	.72		
5. dedicates funds to conduct an evaluation.		.86	
6. dedicates funds to ensure ongoing evaluation.		.97	
7. has resources in place to undertake evaluation on an ongoing basis.		.84	
8. has the commitment from external stakeholders to ensure evaluation sustainability.		.61	
The leadership of my organization...			
9. builds ideas in collaboration with staff members.			.79
10. resolves inter-personal conflicts in a positive manner.			.83
11. celebrates staff members' achievements.			.78
12. promotes evaluative thinking.			.70

Note. α = coefficient alpha.

Structural Equation Modeling

An SEM analysis was used to test the relationships among the individual and organizational EC factors, contributing validity evidence regarding relationships to other variables and internal structure, which addresses measurement standard 1.16 (AERA et al., 2014). The results of this analysis support the inferences generated about individual and organizational EC within the early childhood field. The model includes five latent factors that were used to predict Individual Motivation to Conduct and Use evaluations. The latent factors include Organizational Culture, Organizational Leadership, Organizational Commitment to Evaluation, Individual Attitude Toward Evaluation, Individual Evaluation Skills, and Individual Motivation to Conduct and Use Evaluations.

Initial SEM analysis involved using the factors and all of the items defining them. The results indicated that residuals for several items needed to be correlated in order to achieve an acceptable model-data fit. This approach has a disadvantage because it may yield a model that is highly specific to the present study's sample and difficult to replicate in other contexts. To improve the replicability of the model, a more conservative approach was used, which involved measuring each EC factor by three items with the highest factor loading (Hayduk & Littvay, 2012). Another advantage of using three items to measure each factor is to achieve a stable structure across replications and a parsimonious model (Hayduk & Littvay, 2012; Little, Cunningham, Shahar, & Widaman, 2002). The findings using factors measured by three items indicated a better model-data fit.

The chi-square test of model fit, χ^2 (128, $N = 329$) = 216.9, $p < .001$, was significant, indicating a poor fit. However, the goodness-of-fit criteria, RMSEA = .046 (90% CI [.035, .056]), SRMR = 0.064, CFI = .958, TLI = .95, indicated a good model fit with the data.

Additionally, all direct and indirect paths were statistically significant. Individual Motivation to Conduct and Use Evaluations was positively predicted by the latent variables Organizational Culture, Organizational Leadership, Organizational Commitment Toward Evaluation, and Individual Evaluation Skills and Individual Attitude Toward Evaluation. These relationships explained 49% of the variance in Individual Motivation to Conduct and Use Evaluation. Figure 7 presents the standardized factor loadings and parameter estimates for the paths.
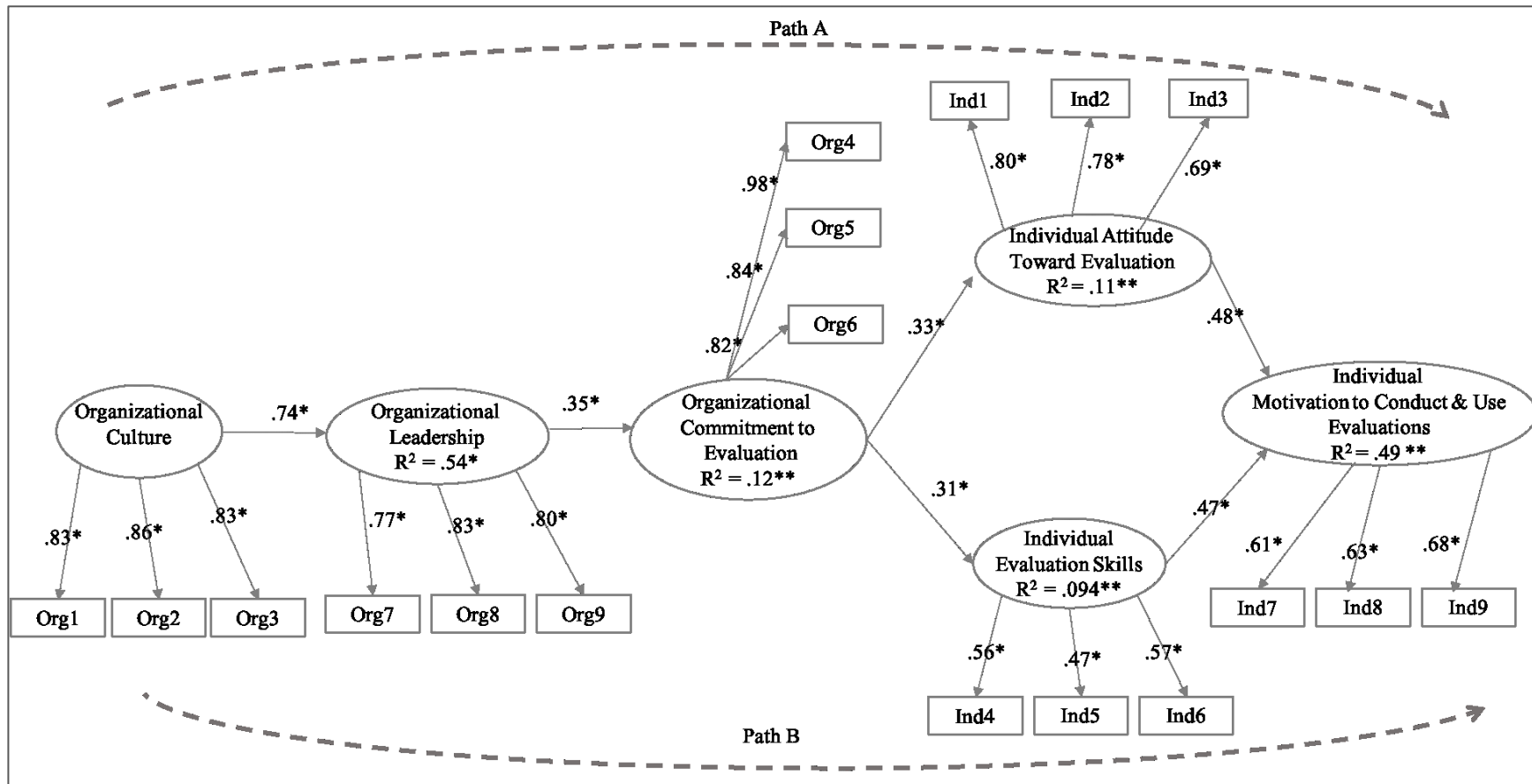
Figure 7. Standardized structural equation model results for the evaluation capacity framework.
*p < .001. **p < .05.
N=329.

Furthermore, the SEM results indicated that Organizational Culture strongly predicted a positive relationship with Organizational Leadership, $\beta = .74$, $p < .001$, $R^2 = .54$. Organizational Leadership in turn moderately predicted a positive level of Organizational Commitment to Evaluation, $\beta = .35$, $p < .001$, $R^2 = .12$. Organizational Commitment to Evaluation also positively and similarly predicted Individual Attitude Toward Evaluation, $\beta = .33$, $p < .05$, $R^2 = .11$, and Individual Evaluation Skills, $\beta = .31$, $p < .05$, $R^2 = .094$. In turn, Individual Attitude Toward Evaluation, $\beta = .48$, $p < .001$, and Individual Evaluation Skills, $\beta = .47$, $p < .001$, strongly and positively predicted Individual Motivation to Conduct and Use Evaluations ($R^2 = .49$). Individual Attitude Toward Evaluation and Individual Evaluation Skills had similar path coefficients, which means that the two factors were able to predict Individual Motivation equally well.

The indirect paths from Organizational Culture to Individual Motivation to Conduct and Use Evaluations were tested with 1000 bootstraps and a bias corrected 95% CI (Preacher & Hayes, 2008). There were two significant indirect paths, as shown in Figure 7. The first indirect effect, Path A, revealed that Organizational Culture indirectly predicted Individual Motivation through Organizational Leadership, Organizational Commitment to Evaluation, and Individual Attitude Toward Evaluation, $\beta = .037$, 95% CI [.021, .106]. Similarly, the second indirect effect, Path B, indicated that Organizational Culture indirectly predicted Individual Motivation through Organizational Leadership, Organizational Commitment to Evaluation, and Individual Evaluation Skills, $\beta = .041$, 95% CI [.028, .108]. The results of the indirect effects can be interpreted as showing that a positive change in Organizational Culture is associated with an increase in Individual Motivation via the prior estimated direct effects.

An additional analysis was conducted to test an alternative SEM differing in one way from the previous model: The predictive relationship between Organizational Leadership and

Organizational Culture was reversed. Similar to the first model, the goodness-of-fit criteria indicated a good model fit with the data, and all paths were statistically significant, RMSEA = .046 (90% CI [.035, .056]), SRMR = 0.064, CFI = .96; TLI = .95. In this model, Organizational Leadership positively predicted Organizational Culture, β = .74, p < .001, R^2 = .54, and Organizational Commitment to Evaluation, β = .36, p < .05, R^2 = .13, while the latter in turn predicted Individual's Attitude Toward Evaluation, β = .33, p < .001, R^2 = .11, and Individual Evaluation Skills, β =.31, p < .001, R^2 = .09. Finally, each of the last two factors predicted Motivation to Conduct and Use Evaluations, β = .48, p < .001, β = .46, p < .001, R^2 = .49. One difference in this model is that there was no connection between Organizational Culture and Individual Motivation to Conduct and Use Evaluations. This finding may be explained by the influence of time on the relationship between leaders and the organizational culture. Leaders may be influenced by the organizational when they initially join an organization, but overtime they may begin to shape the culture. This relationship was not tested given that the data are not longitudinal.

Multi-Group Path Analysis

Finally, I used a multi-group path analysis to examine whether the path coefficients differ across respondents' professional roles, that is, across leaders and non-leaders. The chi-square difference test for the multi-group path, χ^2(6, N = 329) = 12.592, p < .05, was not significant, indicating that respondents' professional roles do not significantly influence the paths from Organizational Culture to Individual Motivation to Conduct and Use Evaluations.

Chapter 5. Discussion and Conclusions

In this chapter, I discuss the study findings and advance conclusions. To that end, I revisit the guiding research questions and discuss the theoretical understandings and psychometric properties inferred from the data generated through the administration of the EC instrument to address the study's research questions. Finally, I describe the study's theoretical, methodological, and practical implications for addressing the pressing need for empirically driven investigations of EC (King, in press). The study was guided by three research questions:

1. What are the underlying EC factors in the early childhood field?

2. To what extent do the sample data support the conceptual EC model in the early childhood development field?

3. Do professional roles influence any of the paths connecting the individual and organizational EC constructs? If so, how?

To address the first question, which focuses on operationalizing EC, the individual and organizational factors were examined using confirmatory factor analysis (CFA). For the second question concerning the hypothesized EC framework, the factor relationships were examined using structural equation modeling (SEM). To answer the third question about the influence of respondents' professional roles on the factor relationships, a multi-group path analysis was conducted. To address these questions, I discuss the findings and implications from these data procedures, this chapter is organized in four parts: (1) a discussion of the findings, (2) a description of the implications, (3) a delineation of the study's limitations, (4) a presentation of future research directions, and (5) end with concluding thoughts.

Discussion of Findings

The discussion of this study's findings is organized by the three research questions. In order to operationalize EC, I first discuss the validity evidence for the multidimensionality of the individual and organizational EC factor structures, fulfilling measurement standards 1.1, 1.11, and 2.3 (AERA et al., 2014). To investigate the hypothesized EC framework, I next discuss the linear relationships between and among the individual and organizational EC factors, adhering to Standard 1.16 (AERA et al., 2014). Finally, to examine potential influences on the model, I consider the non-significant findings of the influence of professional roles on the factor relationships.

Evaluation capacity as a multidimensional construct. This study advances an empirically-based factor structure for organizational and individual EC within the context of the early childhood field. The validity evidence generated by CFA provided consistent support for the multidimensionality of a three-factor organizational EC structure comprising Organizational Culture, Organizational Leadership, and Organizational Commitment to Evaluation; and individual EC structure comprising Individual Attitude Toward Evaluation, Individual Evaluation Skills, and Individual Motivation to Conduct and Use Evaluations. As summarized in Table 15, empirical support is provided by the interpretation of acceptable fit across three sources of measurement evidence for the individual and organizational factor structures. Specifically, for each of the three measurement sources, the table summarizes (a) the recommended ranges for each measurement index, (b) the findings associated with each index for the organizational and individual EC factor structures, and (c) an assessment of fit (i.e., good or acceptable) relative to the recommended range.

Three trends, which apply to both individual and organizational EC, are evident from reading Table 15. First, the goodness of fit indices (i.e., RMSEA, SRMR, CFI, and TLI; Brown, 2006; Hu & Bentler, 1998) fall within the recommended ranges and thus indicate a good fit with the data for each of the factor structures. Second, the coefficient alpha of all factors exceeds .7, indicating that the factors demonstrate strong reliability (Cronbach, 1951; Cortina, 1993). Third, the inter-factor correlations fall within the recommended range (i.e., ≤ .8; Brown, 2006; Cohen, West, & Aiken, 2014). It is interesting to note that while the level of the correlation between Organizational Culture and Organizational Leadership is somewhat high (see Table 15), it still falls within the recommended range. Furthermore, the relationship between these two factors is consistent with the theoretical understanding of their relation (Burns, Kotrba, & Denison, 2013).

Together, these findings suggest that the three organizational and three individual EC factors measure related but distinct EC dimensions, which provide strong validity evidence related to internal structure. Given the absence of an established instrument measuring EC, the discussion in each of the following subsections focuses on examining the features defining each individual and organizational factor in the present study with respect to those measured by comparable instruments. The discussion also draws on other literature in the field of EC and social psychology to provide evidence supporting the operationalization used in the present study.

Table 15

Validity Evidence for the Multidimensionality of the Individual and Organizational Evaluation Capacity Factor Structures

Source of measurement evidence	Organizational (Org) Factor Structure		Individual (Ind) Factor Structure	
Recommended Ranges	Findings	Fit	Findings	Fit
Goodness of Fit Indices				
RMSEA < .08	.055 (90% CI [.040–.069])	good	.060 (90% CI [.049, .071])	good
SRMR < .08	.057	good	.050	good
TLI > .90	.95	good	.90	good
CFI > .90	.96	good	.92	good
Coefficient alpha (α)				
α >.70	.88 – Org Commitment to Evaluation	good	.85 – Ind Attitude	good
	.88 – Org Culture			
	.86 – Org Leadership	good	.86 – Ind Evaluation Skills	good
		good	.73 – Ind Motivation	good
Inter-factor Correlations				
R ≤ .80	.5 – Org Commitment to Evaluation & Org Culture	good	.24 – Ind Attitude & Ind Evaluation Skills	good
	.5 – Org Commitment to Evaluation & Org Leadership	good	.68 – Ind Attitude & Ind Motivation	good
	.8 – Org Culture & Org Leadership	acceptable	.32 – Ind Evaluation Skills & Ind Motivation	good

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; Adjusted BIC = Bayesian information criterion adjusted for model overfitting.

Organizational evaluation capacity factor structure. The empirical operationalization of organizational EC shares both similarities and differences with other comparable EC instruments. As is evident in Table 16, which situates the operationalization of the present study's EC instrument with respect to four other instruments assessing similar EC factors, there are similarities and differences related to the names of various factors and the features they measure. The table reveals that overall, the instrument proposed by Taylor-Ritzler and colleagues (2013) is the most similar to the present study's instrument. Two of that study's factors, Resources and Learning Climate, share number of features with the present study's factors Organizational Commitment to Evaluation and Organizational Culture. In addition, Bourgeois and Cousins (2013) define two of their components, Budget and External, based on features similar to those defining Organizational Commitment to Evaluation.

It is also noteworthy that while all instruments operationalize organizational EC as multidimensional and sometimes use similar terminology to name the factors, the present study's operationalization has noticeable differences from the others. For example, the present study operationalizes Organizational Leadership differently than the relevant instruments, none of which include any of the features defining this factor in this study. In addition, only one other instrument (Taylor-Ritzler et al., 2013) includes a factor that is comparable to the present study's Organizational Culture. In light of these contrasts, it will be useful to discuss the features defining each of the three organizational EC factors in this study's EC instrument in relation to the other relevant literature to support the present study's operationalization.

Table 16

Study Operationalization of Organizational Evaluation Capacity Constructs

	Instruments			
Factor labels Key features	Taylor-Ritzler et al. (2013)	Cheng & King (2016)	Bourgeois & Cousins (2013)	Fierro (2012)
Organizational Commitment to Evaluation	*Resources*	*Evaluation Infrastructure*	*Budget, External Supports*	*Resources*
Dedicate funding to evaluation	✓	✓	✓	✓
Provides ongoing resources for evaluation	✓	--	✓	--
Has commitment from external stakeholders for evaluation	✓	--	✓	--
Organizational Culture	*Learning Climate*	*Evaluation Culture*	No Equivalent Factor	No Equivalent Factor
Open to new ideas	✓	--	--	--
Involves staff in long-term plans	✓	--	--	--
Provides staff with opportunities for reflection	✓	--	--	--
Involvement in reviewing mission, vision, and values	--	--	--	--
Organizational Leadership	*Leadership*	No Equivalent Factor	*Leadership*	*Supervisors*
Collaborates with staff	--	--	--	--
Resolves conflicts positively	--	--	--	--
Celebrates staff's achievements	--	--	--	--
Promotes evaluative thinking	--	--	--	--

Note. ✓: Feature addressed by the instrument. --: Feature not addressed by the instrument.

The first factor identified in the present study, Organizational Commitment to Evaluation, is defined by three features that focus on whether an organization dedicates funding to evaluation, provides resources for evaluation on an ongoing basis, and has a commitment from external stakeholders to undertake the evaluation. These three features are also used to define similar concepts in two other instruments (Taylor-Ritzler et al., 2013; Bourgeois & Cousins, 2013). In contrast, the instruments developed by Cheng and King (2016) and Fierro (2012) conceptualize this factor very differently. Their equivalent factors, *Evaluation Infrastructure* and *Resources*, share only one feature related to funding for evaluation. The EC literature supports a more expansive view of the level of organizational investment in evaluation that requires more than financial and technical resources (Harris, 2017; King, in press). Specifically, stakeholders' contribution to evaluation reflects a strong buy-in that is required if an organization aims to conduct evaluations that are highly useful (Patton, 2011a). It may also indicate the likelihood of the long-term use of evaluation (e.g., Alkin & King, 2017; Bourgeois & Cousins, 2013; Patton, 2011a).

The second factor in the present study's EC instrument, Organizational Culture, is defined in terms of openness to ideas, staff's involvement in long-term planning and review of the organizational mission, vision, and values, and the availability of opportunities for reflection. This factor shares conceptual similarities with a factor in only one other study, namely, Learning Climate in the instrument proposed by Taylor-Ritzler and colleagues (2013). Both factors focus on openness to new ideas, staff involvement in long-term plans, and the practice of reflection. However, the present study differs insofar as it classifies the involvement of staff in reviewing the organizational mission as an aspect of Organizational Culture, not of Leadership. While it is widely recognized that involving employees in this process is essential, scholars are divided on

how to classify this. Cousins and colleagues (2014) and Preskill and Torres (2001) regard it as an aspect of culture, while Labin (2014) views it as a characteristic of leadership. Furthermore, unlike the present study, Taylor-Ritzler and colleagues include attributes related to staff's ability to encourage their peers or managers to use evaluations.

This study's operationalization of Organizational Culture differs significantly from that of Cheng and King (2016), who use similar terminology but include none of the same features in their operationalization. Instead, they focus on respondents' perceptions regarding evaluation and their willingness to learn about and use evaluations. In the present study, these attributes were included to assess individuals' attitudes because they indicate individual dispositions toward conducting and using evaluations (Fishbein & Ajzen, 2011). As in the present study, the features Cheng and King (2016) use to assess Organizational Culture are consistent with their cited importance in the EC literature, as they are key ingredients in building a culture that is favourable to evaluation (McCoy et al., 2013).

Finally, the present study operationalizes Organizational Leadership in terms of whether the leader is inclined to collaborate with staff, resolve conflicts positively, celebrate achievements, and promote reflective thinking. This operationalization is different from those of Taylor-Ritzler and colleagues (2013), Bourgeois and Cousins (2013), and Fierro (2012) which, although they use similar terminology to name their factors, share none of the present operationalization's features. Unlike the present study, Taylor-Ritzler and colleagues (2013) operationalize this factor in terms of the quality of the tasks completed by organizational leaders, such as clearly communicating goals and the organizational mission as well as making plans to accomplish the goals. Bourgeois and Cousins (2013) focus on whether the leaders possess skills related to evaluation and management. The attributes that the present study includes in assessing

this construct are those that have been reported by other studies to be valuable for conducting and using evaluations (Cousins, Goh, Clark, & Lee, 2004; Labin et al., 2012). For example, leaders who encourage collaboration and promote evaluative thinking are more likely to discuss evaluation findings in meetings and request input from staff (Burns et al., 2013). By offering an empirically informed operationalization of this construct that is consistent with the theoretical understanding of leadership, the present study advances the measurement of this factor.

Individual evaluation capacity factor structure. The operationalization of individual EC in the present study shares relatively few similarities with the other instruments measuring comparable EC factors. Table 17, which situates the factor structure of the present study with respect to three instruments assessing comparable factors reveals trends similar to those observed for the organizational factors. First, the present study's operationalization is somewhat comparable to that of Taylor-Ritzler and colleagues (2013); the factor Individual Attitude Toward Evaluation, shares several attributes with Taylor-Ritzler and colleagues' Awareness factor. The present study's operationalization shares only limited features with the other two instruments (Cheng & King, 20016; Fierro, 2012), even in the case of factors that are labelled using similar terminology. As in the case of organizational factors, it is useful to provide support for my approach by discussing the features defining each of the three individual EC factors in comparison to those assessed by the other instruments and in relation to other relevant literature.

Table 17

Study Operationalization of Individual Evaluation Capacity Constructs

Factor labels 　Key features	Instruments Taylor-Ritzler et al. (2013)	Fierro (2012)	Cheng & King (2016)
Individual Skills	*Competence*	*Knowledge and Skill*	*Human Resources*
Able to hire an external evaluator	--	--	--
Able to oversee an external evaluator	--	--	--
Able to use evaluation findings	--	--	--
Individual Attitude	*Awareness*	*Attitude*	No Equivalent Factor
Usefulness of evaluation for achieving effectiveness	✓	✓	--
Usefulness of evaluation to stakeholders	✓	--	--
Participation in evaluation tasks is one's responsibility	✓	--	--
Individual Motivation	*Motivation*	*Motivation*	No Equivalent Factor
Funding (internal and external) for evaluation is justified	--	✓	--
Training (credentialed and non-credentialed) in evaluation is important (in general and for staff)	✓	--	--
Open to adopting new ideas based on evaluation findings	--	--	--

Note. ✓: Feature addressed by the instrument. --: Feature not addressed by the instrument.

　　　　This study operationalizes the first factor, Individual Evaluation Skills, as individuals'
ability to hire or oversee an external evaluator and to use evaluation findings in their work.
Unlike the present EC instrument, the instruments proposed by Taylor-Ritzler and colleagues
(2013), Cheng and King (2016), and Fierro (2012) consider different features to define

comparable constructs (see Table 17). While the Individual Evaluation Skills factor measures evaluation expertise that is somewhat broader than those assessed by the other instruments, the present EC instrument also includes another section that assesses specific evaluation skills (e.g., creating evaluation plans, collecting quantitative data, and conducting data analysis). The purpose of this section was to inform future skill development strategies. Measuring evaluation skills is challenging because organizational members use different skills depending on their professional role or the task they are involved in (Galport & Azzam, 2017). To address this challenge, the present EC instrument generated data about respondents' perceived skill level and the importance of evaluation to their work. Thus, while assessing specific evaluation skills similarly to the other instruments, the present instrument's operationalization of EC includes general individual evaluation skills that may be relevant to many individuals within an organization.

The second factor, Individual Attitude, assesses respondents' perception of the usefulness of evaluation for achieving effectiveness within the organization, the usefulness of evaluation for stakeholders, and their sense of responsibility for evaluating their programs. These three features are also used to assess a comparable factor, Awareness, in Taylor-Ritzler and colleagues (2013). In contrast, Fierro's (2012) instrument uses only one of these features—the benefit of evaluation for improving effectiveness—to define a comparable concept, Attitudes. The operationalization of the Individual Attitude factor in the present study is consistent with a belief-based approach in which beliefs are used to indicate respondents' inclination to participate in a behaviour (e.g., conducting and using evaluations; Fishbein & Ajzen, 2011; Thurstone, 1928).

Lastly, Individual Motivation is defined by features related to respondents' inclination to invest time and money in improving their skills to conduct and use evaluations as well as their

willingness to support their colleagues in taking part in the evaluation. While the other instruments use similar terminology, the features they use to define their concepts differ considerably. Fierro (2012), for example, focuses on the reasons for conducting an evaluation (e.g., to be accountable, improve the program), while Taylor-Ritzler and colleagues (2013) ask direct questions about whether respondents are motivated to conduct and use evaluations. These differences in defining motivation reflect the varied ways motivation has been discussed in the theoretical EC literature (Labin et al., 2012; Preskill & Boyle, 2008). In the present study, Individual Motivation represents a proxy for measuring the extent to which individuals may adopt a behaviour, such as conducting and using evaluation. Assessing respondents' motivation to adopt a behaviour may be best captured by inquiring about their willingness to invest their limited resources (e.g., time and money) to conduct an evaluation and use evaluation findings (Ajzen, 2005; Burke et al., 2008; Fishbein & Ajzen, 2011).

In sum, the validity evidence substantiated the multidimensionality of the individual and organizational EC constructs, indicating that the six EC factors are related but represent distinct dimensions. Comparing the operationalization of EC factors to that of other comparable instruments reveals both similarities and differences. The differences, which include the use of different features to operationalize the same construct, draw attention to the need for an empirical understanding of EC, which this study begins to address.

Pathways for predicting evaluation capacity. This study advances EC conceptualization by testing an empirically based EC framework within the early childhood field. The validity evidence generated by SEM, related to internal structure and relations to other variable, provided consistent support for this framework. The goodness-of-fit indices RMSEA < .05, SRMR < .08, CFI > .90, and TLI > .90 indicated that the model had a good fit with the data

(Brown, 2006; Hu & Bentler, 1998). These results provide evidence in support of the relationships between and among the individual and organizational EC components tested by the SEM. In addition, the SEM's bootstrap test was significant at a 95% confidence interval, which provided statistical evidence in support of the indirect paths between the organizational and individual EC factors. These relationships can be better understood by reflecting on their direction and magnitude, which are measured by the path coefficients (r) and the coefficients of determination (R^2), respectively. Figure 8 provides a visual illustration of the direct and indirect effects predicted by relationships among the EC factors. To integrate the SEM findings, this section discusses the direction (r) and magnitude (R^2) for each of three sets of relationships: relationships among the organizational EC factors, relationships among the individual EC factors, and the direct and indirect relationships between individual and organizational EC factors. The discussion in each subsection draws on studies by EC scholars and others to parse out these relationships, expanding and detailing potential EC processes. Thus, this study contributes a new way to approach EC building, in which organizations can make informed and data-driven decisions.
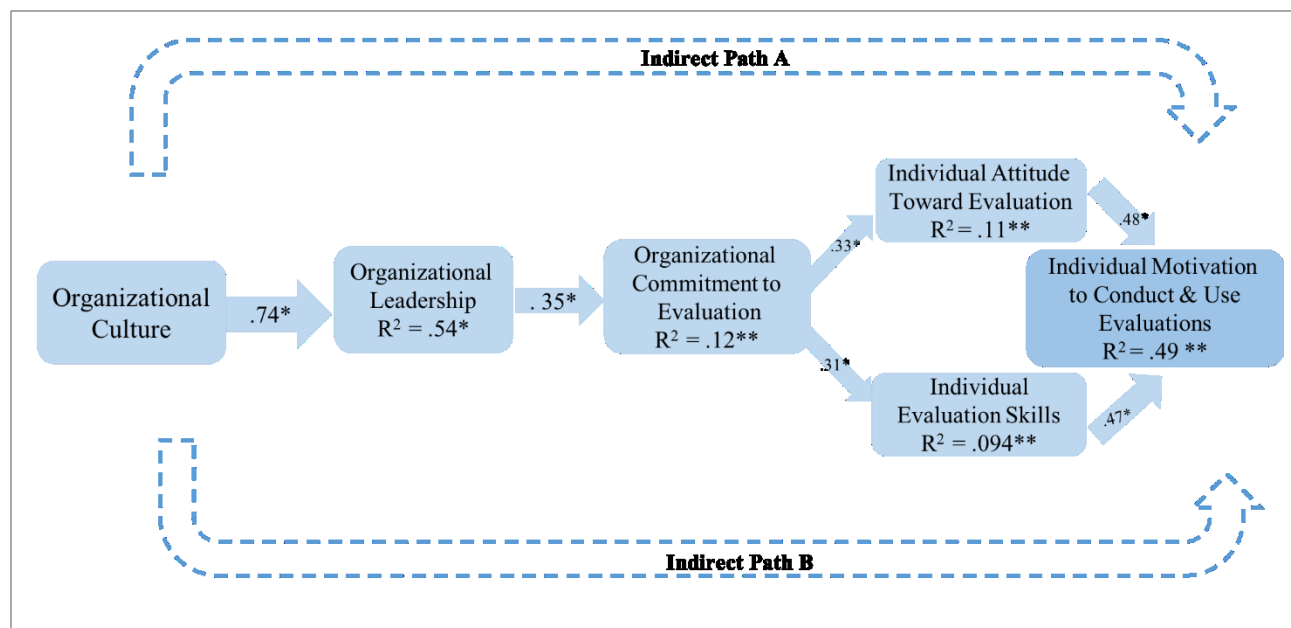
Figure 8. Relationships among the individual and organizational evaluation capacity factors based on the structural equation modeling.

*p < .05. **p < .001.

Relationships among organizational factors. The SEM findings highlighted a set of robust predictive relationships among the organizational factors. In the first of these, Organizational Culture positively predicted Organizational Leadership ($r = .74$; $R^2 = .54$). In the second predictive relationship, with a lesser magnitude, Organizational Leadership influenced Organizational Commitment to Evaluation ($r = .35$; $R^2 = .12$). While the magnitude of the second relationship explained only 12% of the variance in Organizational Commitment to Evaluation, it is statistically significant, with a 99% confidence level (Preacher & Hayes, 2008). These findings appear to be consistent with the understanding of these relationships by other scholars from perspectives grounded in theory and practice.

The strength of the first relationship, between Organizational Culture and Organizational Leadership, appears to be consistent with theorized expectations. The literature has identified a firm theoretical basis for the relationship between the leaders of an organization and its culture (Burke, 2017; Lamm & Israel, 2011; Preskill & Boyle, 2008a). It is noteworthy that, in addition to asserting that leaders shape the culture of their organizations, the literature also observes that they may also be shaped by it (e.g., Burke et al., 2008; Burke & Litwin, 1992). However, previous studies have tended to focus on the first direction: organizational leadership's influence on culture.

While the SEM results in the present study provided statistical evidence in support of a relationship between Organizational Culture and Organizational Leadership in both directions, the findings provide stronger support for the first direction: culture to leadership. One reason, interestingly, is that Organizational Leadership directly predicted Organizational Commitment to Evaluation regardless of the direction, but this was not the case for Organizational Culture. These findings may mean that leaders who exhibit characteristics conducive to evaluation may positively influence their organization's overall commitment to evaluation. Evidence from practice supports this conclusion (Goh, Cousins, & Elliott, 2006; King, 2007; Lamm & Israel, 2011). For example, King and Volkov (2005) found that the process of EC building was stronger in organizations in which the leadership was committed to evaluation and envisioned how to use evaluation results to improve their programs and enhance learning.

Relationships among individual factors. The second finding, which involved the significant direct relationships among the individual factors, also provided strong support for the predictive relationships among the individual EC factors. The SEM revealed that Individual Attitude Toward Evaluation ($r = .48$) and Individual Evaluation Skills ($r = .47$) exerted a similar

influence on Individual Motivation to Conduct and Use Evaluations. They explained 23% and 22% of the variance, respectively, at a 99% confidence level (Preacher & Hayes, 2008). These findings indicate that in such cases, where members of organizations have a positive attitude and are skilled in evaluation, it is reasonable to expect that they will exhibit high motivation to conduct and use evaluations. Conversely, organizations may face challenges when the opposite is true. When individuals in an organization have both negative attitudes and low skill levels, for example, it may not be sufficient to focus on improving only one of these.

The dynamics may be best understood from the perspective of motivation theories, particularly the theory of planned behaviour (Ajzen, 1991; Ajzen & Fishbein, 1977; Fishbein & Ajzen, 2011), which aims to predict an individual's motivation to engage in a behaviour. In the case of EC, the desired behaviour is to have organizational members who are motivated to conduct evaluations and use their results. This theory explains individuals' intentions or motivations based on their beliefs about the value of an action and whether they perceive that they are able to engage in it or have control over it (Ajzen, 1991). Ajzen's and Fishbein's work offers insights that are useful for explaining the relationship between the EC factors examined in this study. Ajzen explains, for example, that attitude is an important determinant of whether individuals will be motivated to engage in a behaviour (Ajzen, 2005), and while Ajzen's theory does not refer to skills, this factor can be presumed to partially correspond to individuals' perceived ability to engage in a behavior, such as being actively involved in conducting an evaluation. Possessing the needed evaluation skills will likely give organizational members confidence in their ability to conduct and use evaluations.

Direct and indirect relationships connecting individual and organizational factors. The

third set of relationships provided an important contribution to understanding the specific

interactions between individuals and organizations that are revealed in the direct and indirect

predictive relationships between their corresponding factors. Regarding direct relationships, the

SEM indicated that Individuals' Attitude Toward Evaluation ($r = .33$, $R^2 =. 11$) and Individual

Evaluation Skills ($r = .31$, $R^2 = .09$) were directly influenced by Organizational Commitment to

Evaluation. These two relationships had the same direction and magnitude, which suggests that

Organizational Commitment to Evaluation may equally influence Individuals' Attitude Toward

Evaluation and Individual Evaluation Skill levels. While the variance explained by

Organizational Commitment to Evaluation may appear small, it is significant, with a 99%

confidence level (Preacher & Hayes, 2008). Regarding indirect relationships, bootstrap analysis

of the SEM at a 95% confidence level (Preacher & Hayes, 2008) revealed two statistically

significant indirect effects (Paths A and B in Figure 8) in which Organizational Culture indirectly

predicts Individual Motivation to Conduct and Use Evaluations, a relationship that is mediated

by the other individual and organizational EC factors. This finding indicates that each factor in

the path plays a crucial role in the indirect effects.

These direct and indirect relationships between the EC factors are largely consistent with

Ajzen's (1991) theory of planned behaviour. In particular, the factors address three features of

his theory, which postulates that (a) a positive attitude, (b) favourable subjective norms, and (c) a

high degree of perceived behavioural control are the best predictors of individuals' positive

motivation to engage in a given behaviour. Organizational Culture represents the social norm

that Ajzen articulates as the social pressure to perform or not perform the intended behaviour.

Similarly, when the organizational culture is favourable to evaluation, it may lead to a cascade of

other positive influences. This norm, which is mediated via organizational leaders, may be reflected in the degree to which organizations appear committed to evaluation in terms of the availability of resources and supports. Their commitment (or the lack thereof) influences individuals' perceived control over their ability to conduct and use evaluations. An individual's perception of a high level of control may be reflected in the individual's positive attitude toward evaluation and the evaluation skills he or she possesses, while a low level of perceived control may lead to the opposite.

The two indirect paths of the empirical EC framework highlight the importance of viewing EC from a systems perspective. The nature of the direct and indirect relationships along these paths suggests that individual and organizational EC factors mediate the degree to which organizational members are motivated to conduct and use evaluations. When one of the EC factors is not functioning optimally, it may negatively affect the other EC factors—particularly organizational members' motivation to conduct and use evaluations, which is the desired outcome of the EC process. As such, this study's empirical framework highlights the importance of treating the conceptualization, measurement, and practice of EC from a systems perspective, which is also consistent with the organizational change theory developed by Burke (2008, 2017). Burke's theory emphasizes that it is important to approach change (e.g., enhanced motivation to conduct and use evaluations) by understanding the entire system, even if efforts target only one component at a time. The present study contributes to an integrated understanding of the individual and organizational EC constructs and, in doing this, begins to address the recent calls for a more precise conceptualization and measurement of EC (Cheng & King, 2016; Gagnon et al., 2018; King, in press).

In sum, the three sets of EC relationships provide empirical evidence that elucidates the nature of the interaction between the individual and organizational EC factors. Compared to the existing literature, this study offers an integrative understanding that provides more specificity by examining and discussing the directionality and magnitude of these relationships. In so doing, this study draws attention to the importance of approaching EC conceptualization, measurement, and practice from a systems perspective in which EC factors are examined as a collective sum and not merely as disjointed parts.

Influence of respondents' professional roles on evaluation capacity pathways. Addressing the third question entailed empirically assessing the influence of respondents' professional roles on the two EC pathways—that is, investigating whether the direct and indirect relationships between and among the EC factors might differ between groups with different professional roles. The multi-group path analysis revealed that respondents' professional roles did not have a statistically significant influence on the EC pathways. It is difficult to compare this finding to the findings of other studies, because other studies have not assessed the influence of the respondents' role on the relationships among EC factors. Instead, the few existing studies that consider the respondents' roles (i.e., Cousins et al., 2008; Fierro & Christie, 2017) have investigated whether responses differed based on the respondents' role. In these descriptive studies, the authors found that managers reported a more positive assessment of the organization's existing EC than those who were directly involved with evaluations, such as front-line employees. These studies' findings are consistent with what other scholars (e.g., Taylor-Ritzler et al., 2013) have noted as they emphasized the importance of gathering data from diverse respondents representing the wide-ranging organizational involvement in evaluation.

While the present study did not statistically assess mean differences among respondents with different roles because the data did not adhere to the required assumptions, descriptive statistics showed some slight discrepancy between respondents in leadership roles and those who were not in leadership roles. From a measurement perspective, it is of consequence that the findings suggest that the relationships among the EC factors appear to be stable regardless of respondents' professional roles. In sum, the absence of an influence of roles in this study reveals that the factor structure and the paths among the factors appear to be measured similarly regardless of the respondents' professional roles. For this study's sample, this may have validity implications supporting the internal structure of the EC instrument and providing evidence that the EC instrument can assess the individual and organizational EC constructs equally well for all respondents. Indeed, this finding does not argue against the importance of using the EC instrument with a representative sample to accurately assess the existing level of EC across the individual and organizational factors.

Study Conclusions

This study has three main implications for advancing theory, methods, and practice related to EC. First, the study advances EC theory based on an empirically developed framework that specifies individual and organizational EC factors, the relationships among them, and the magnitude of those relationships within the context of early childhood. The study also advances a methodology for embedding the conceptualization and measurement of EC in the context in which it will be used to generate conclusions about EC that are relevant and useful. Finally, it applies an evidence-based approach to understanding individual and organizational EC. Taken together, these contributions begin to address the call from Preskill (2013), as well as other scholars and practitioners, to focus our scholarly energy on how to address the "hard stuff"—that

is, how to undertake the complex tasks of building EC and evaluate its resulting impacts

systematically:

> I believe it is now time for evaluators to [translate] our collective wisdom into
>
> better and deeper practice. In particular, we need to focus on ensuring that our ECB
>
> [EC] efforts make a difference—that they reach the right people, that they are
>
> designed and implemented for learning transfer, and that ECB [EC] activities are
>
> evaluated for their effects, influence, and impact. (p. 2)

By developing a comprehensive theory, employing rigorous methods, and enhancing evidence-

based practice, this study takes an important step toward achieving these goals in a way that may

benefit scholars and practitioners both within and outside of the field of early childhood

development. This section highlights the three implications of this study, outlines two study

limitations, and delineates two major lines of inquiry for future directions. I conclude by

reflecting on the purpose and contribution of this dissertation study.

Theoretical implications. This study advances a comprehensive empirical EC

framework that is relevant to the context of the early childhood field. This new framework

contributes to EC theory by specifying the magnitude and directionality of the influences among

individual and organizational EC factors and by providing evidence that supports direct and

indirect paths linking individual and organizational factors. These complementary features of the

framework—its specificity and its comprehensiveness—are crucial for advancing our theoretical

understanding of EC. Furthermore, the instrument provides a way to empirically assess the EC

framework as well as investigate new relationships that may be relevant in different contexts.

This study's empirical approach has already yielded valuable insights into the

relationships between and among EC factors. In assessing the framework, for instance, this study

has identified and affirmed that Individual Motivation to Conduct and Use Evaluations was an outcome of the EC process in the context of the early childhood field. This behavioural outcome serves as a proxy for predicting individuals' intention to carry out evaluations and use their results. To be sure, identifying motivation as an important EC component is not new (e.g., Cheng & King, 2016; Fierro, 2012; Labin et al., 2012; Taylor-Ritzler et al., 2013). Nonetheless, this study makes a key contribution by empirically (and conceptually) identifying individuals' motivation as an immediate outcome that is influenced directly by organizational members' evaluation skills and their attitude toward evaluation and indirectly by an organization's culture and leadership and its level of commitment to evaluation. Bearing in mind that components making up the EC system may be fluid; other scholars may investigate those and other EC components and their relationships in different contexts.

By explicitly identifying these EC factors and empirically evaluating the relationships among them, this study has made theoretical assumptions explicit. In so doing, it may allow other scholars to evaluate or extend them. For example, they may consider the degree to which context influences the operationalization of the EC factors. They may also explore additional factors that are relevant to their studies. It is important to note that this exercise requires scholars to balance the need for rigour and comprehensiveness with the meaningfulness and applicability of findings. To navigate this balance, particularly in the areas of evaluation and EC, it is highly advantageous to integrate theory with practice to identify the critical issues that research evidence can support. This study contributes to "sense making," as Patton called it, by providing a framework with the most critical factors that "deserve priority" (Patton, 2014, p. 239).

Methodological implications. This study adopts a context-relevant approach to developing a conceptual EC framework and instrument. Context is multidimensional, reflecting

sets of conditions that may involve processes and/or sets of behaviours (Ghoshal & Bartlett, 1994). Because of this multidimensionality, it is difficult to assess context as a latent construct. In fact, the context in which organizations exist can influence various individual and organizational practices and cultures, including those that are consequential for the EC process. In this study, the context was considered by involving researchers and practitioners with diverse knowledge of measurement, evaluation and the field of early childhood. Given that this study was part of a larger project, it was possible to involve experts in assessing different aspects of the context early on during the conceptualization and operationalization of EC.

The methodological approach for this study highlights the need for measurement tools that are both useful to practitioners and adhere to appropriate measurement practices. This study navigated this balance by including experts early in the process and by using strategies that improved the perceived usefulness of the instrument. For example, it was more effective to integrate semi-structured consultations to discuss the conceptualization and operationalization of the construct during established meetings than to use the Delphi method (Landeta, 2006). While using the latter method may have provided a structured approach to reviewing the tools and achieving consensus among the experts, the consultations allowed for discussions and exchanges that highlighted points of conversion as well as points of diversion. Capturing areas of diversion were important to create a conceptualization and operationalization of EC that increased its usefulness to the stakeholders. By representing stakeholders in the process, scholars may enhance their perceptions of the usefulness of an instrument, which may increase their willingness to complete it. This process allowed scholars and practitioners to assess EC with more accuracy, leading to relevant recommendations to participating organizations.

The use of measurement tools, including the EC instrument, should be informed by the purpose of such use, the context in which it is used, and potential consequences of the use. One of the reasons that this study did not attempt to operationalize context was that, in addition to its multidimensional nature, defining it would inevitably lead to over-simplification of a complex construct. While I agree with Daigneault and Jacob (2009) and Blumer (1954) about the importance of operationalization to advancing scholarly understanding of complex social concepts, Patton (2011b) makes the pertinent point that some concepts, such as context, are best left undefined. One reason is that doing so may restrain the "exploratory and innovative processes" (p. 265).

A good way to consolidate measurement rigour with context relevance is to involve potential users (i.e., stakeholders) early in the process. Given that validation is a continual process, practitioners need to be conscious of evaluating the appropriateness of the instruments and the resulting conclusions within the context of their study. This is especially important when the context is significantly different from that for which the instrument was initially created. For example, methods such as Delphi and cognitive interviewing or think-aloud may be more useful tools to use to evaluate the relevance of items in one context over another. Scholars should be guided by their understanding of their stakeholders to inform their decisions about which approaches to use. In so doing, scholars can balance the need to adhere to measurement rigour without being too rigid in applying those principles. Involving stakeholders in the process has other benefits such as improving buy-in and mitigating misuse, which minimizes the negative consequences of using the instrument. Ultimately, the goal is to produce tools that are contextually relevant so that the inferences generated from the measurement tools are valid.

Practical application. This study produced an EC instrument with the capability of providing evidence-based guidance to organizations seeking to improve their individual and organizational EC. The EC practice was the source of inspiration for this study. By assessing the level of EC identified in the individual and organizational factors and understanding the direct and indirect influences of each EC factor, organizations can use this study's findings to invest in initiatives that target their specific EC needs. In this way, organizations can align their efforts with those initiatives that will give them a more significant return on their investment.

Aligning efforts to the most needed areas may be especially valuable for organizations with limited capacity in one or more EC factors. In such cases, it may be beneficial to devise efforts targeting specific factors that will lead to improvement. For example, initially directing resources toward building evaluation skills might not be the best approach if organizational EC is limited. By contrast, by focusing initially on organizational culture and leadership, organizations may be able to develop an EC foundation that leads to greater improvement for the remaining four EC factors (Organizational Investment in Evaluation, Individual Attitude Toward Evaluation, Individual Evaluation Skills, and Individual Motivation to Conduct and Use Evaluation) among their members. Ultimately, the goal of contributing a specific and comprehensive framework and instrument for assessing EC factors is to enable organizations to make informed decisions about which aspects of their EC require improvement and to periodically reassess those factors.

An inescapable challenge when offering to adopt rigours measurement within an applied setting is that practitioners tend to lack statistical and measurement skills. This challenge is especially prevalent among practitioners within community organizations who might be considering using an EC instrument. An effective and relatively less expensive solution to this

problem is to form a community-university partnership. Academics and students can provide technical support, while organizations can provide opportunities to researchers and students to study the social and methodological implications of interest and apply that knowledge to solving the challenges facing organizations. This dissertation study is part of a larger community-university partnership project that sought to build evaluation capacity in the early childhood field (Gokiert et al., 2017). The collaboration between researchers and practitioners should continue to guide our scholarly work to produce results that are relevant and best meet organizations' needs.

Study limitations. Two limitations should be considered with respect to this study. The first limitation relates to the procedures that were used for the initial sample recruitment. The initial recruitment letters invited leaders or managers to attend a discussion forum as part of the overall project and to complete the EC instrument for the purpose of this study. However, the letters did not explicitly ask recipients to distribute the EC instrument to others beyond administrators in their organization. This had unintended consequences, as it may have generated an unequal distribution of respondents, with an overrepresentation of individuals in leadership positions. This may limit the usefulness of the findings to assess learning needs because often individuals' professional roles determine their EC needs (Fierro & Christie, 2017). If the desired outcome is to improve EC for organizational members who work in the front-line, then a sample that does not adequately represent different perspectives may lead to flawed conclusions about strengths and gaps in the individual and organizational EC factors, resulting in initiatives that are not effective. Efforts to mitigate the limitations of the initial sample included sending a second invitation letter about five months later to recruit respondents with different professional roles. Future studies can mitigate this limitation by using stratified random sampling (Blair et al., 2013), which should allow the perspectives of different stakeholders to be considered.

The second limitation relates to the data collection procedures, which were dependent on respondents' self-reports on the EC instrument. As in any study that relies on self-reports, respondents' answers to the items may have been influenced by the social desirability bias, a situation in which respondents answer questions according to what they perceive as acceptable as opposed to what is true (Blair et al., 2013). In the case of assessing EC, responses might be positive or negative depending on what respondents perceive the goals of the assessment to be (Fierro, 2012). For example, positive responses may result from respondents wanting to present their organization in a favourable light, especially if they are in a leadership position. Negative responses, on the other hand, may reflect a belief that reporting a low EC may lead to more investment in this area, such as more funding for training (Fierro, 2017). This bias is an issue because it can contribute to an inaccurate assessment of existing EC, and could lead to an increase in measurement error, weakening the reliability of the results. Efforts to mitigate this limitation included ensuring respondents' anonymity during data collection, data analysis, and dissemination. To mitigate the influence of this bias in this study, one negatively worded question was included to evaluate the consistency of respondents' answers, thus providing some support for the reliability of the findings. Future studies might consider triangulating multiple data sources, which could improve the accuracy of the findings by highlighting areas of convergence and divergence (Latcheva, 2011).

Beyond these methodological limitations, two challenges emerged that were related to the specific context of the early childhood field including (1) the diversity of organizations, and (2) the limited range of studies of EC that are specific to this context. Regarding the first challenge, the early childhood field includes organizations of different sizes specializing in different areas (e.g., childcare, accreditation, learning, health). While this diversity provides opportunities for

supplying a broad range of services that meet clients' needs, it makes it challenging to create a common understanding for EC conceptualization and measurement. Regarding the second challenge, given the dearth of studies that discuss EC conceptualization or measurement within the early childhood field, it was difficult to distinguish the unique characteristics of EC in that context that influence conceptualization, measurement, and practice. To mitigate the impact of this challenge, I consulted with diverse experts on the relevance of the conceptualization and the items created to operationalize the components to the study's context. While it is difficult to capture the varying influences of contexts, it will be important for other scholars to be cognizant of the extent to which they are taking steps to capture the varying aspects of context, especially those that may influence their assessment of EC.

Directions for future research. Two critical directions for future research include examining the relevance of the EC instrument in other contexts, and assessing its ability to capture the impact of EC efforts over time. Given that EC is context specific, the first direction involves evaluating the appropriate use of an EC instrument to assess individual and organizational EC factors in other contexts. This is important because it has the potential to expand our understanding of the contextual characteristics that influence the conceptualization and measurement of EC, and it may help researchers determine whether this study's findings are generalizable. Initially, this could entail replicating the study in organizations within the early childhood field located in a different province. Policies related to the early childhood field, while somewhat regulated by the federal government, normally fall within the provincial and territorial jurisdiction (Langford & Richardson, 2018). The lack of common regulations is reflected in different funding levels and priorities (Ferns & Friendly, 2012), which likely influences the context in which organizations function. Such an investigation might elucidate the degree to

which context influences the conceptualization and measurement of EC in closely related contexts.

Evaluating the relevance of this study's conceptualization and measurement of EC may also require multiple investigations in other fields such as education or health. Using the EC instrument in a new context requires gathering validity evidence to evaluate its relevance to that context (AERA et al., 2014). In addition to testing the EC framework and the factor structure by replicating the CFA and SEM analyses used in the present study, it may also be useful to evaluate the relevance of the content assessed by the instrument using cognitive interviews or focus groups. Such studies would expand our understanding of the contextual characteristics that might influence the generalizability of findings about assessing individual and organizational EC.

The second future research direction involves investigating the degree to which the EC instrument can capture changes in EC over time. As organizations invest time and money to develop their EC, it becomes important to assess the impact that those efforts have on EC building over time. Longitudinal use of the EC instrument would require examining the degree to which the EC instrument captures changes in the individual and organizational EC factors. Such an investigation would require evaluating the factor structure and longitudinal measurement invariance of the EC constructs and their relationships (Schumacker & Lomax, 2012). In this line of inquiry, scholars could investigate other paths of influence among the individual and organizational EC relationships that might be observable over time, such as Organizational Leadership's influence on Organizational Culture. This would require examining validity evidence that can be generated using latent variable growth modeling, allowing scholars to assess changes in individual and organizational EC factors over time (Byrne & Crombie, 2003). An

instrument that allows organizations to track the impact of their EC initiatives over time would help them devise evidence-based strategies that target areas that need improvement and also celebrate their ongoing successes.

Concluding Remarks

This dissertation study contributes to the EC literature in three key ways. First, it provides an empirically-based theoretical framework that conceptualizes EC by specifying the individual and organizational factors and the direction and magnitude of their relationship. Second, it embeds a methodology for generating valid conclusions that are contextually relevant within the early childhood field. Finally, it develops an EC instrument that measures individual and organizational EC factors and assesses their relationships. In so doing, the study offers scholars an illustrative example for use in investigating EC empirically in other settings. It also offers practitioners an EC instrument to guide their efforts in developing individual and organizational EC effectively and efficiently from a systems perspective. The study reflects my belief that scholarly pursuits are strengthened and enriched when we draw on theory, methods, and practice. The benefits of merging these areas can be seen clearly in the areas of evaluation and EC. Building EC within organizations requires thoughtful preparation that may include ongoing collaborations between scholars and practitioners to refine our conceptualization, measurement, and practice. The findings of this study may empower organizations and their members to make choices based on informed decisions that are empirically supported.

References

Abma, T. A. (2006). The social relations of evaluation. In I. F. Shaw, J. C. Greene, & M. M.

    Mark (Eds.), *The Sage handbook of evaluation* (pp. 184–199). Thousand Oaks, CA:

    SAGE.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision

    Processes*, *50*(2), 179–211. Retrieved from https://doi.org/10.1016/0749-5978(91)90020-

    T

Ajzen, I. (2005). *Attitudes, personality, and behavior* (2nd ed.). Berkshire, UK: McGraw-Hill

    Education.

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review

    of empirical research. *Psychological Bulletin*, *84*(5), 888–918.

Ajzen, I., Heilbroner, R. L., Fishbein, M., & Thurow, L. C. (1980). *Understanding attitudes and

    predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Akaike, H. (1980). Likelihood and the Bayes procedure. In J. M. Bernardo, M. H. De Groot, D.

    V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 143–166). Valencia, Spain:

    University Press.

Alberta Government. (2013). *Summing up: What Albertans said about early childhood

    development*. Alberta, Canada. Retrieved from

    http://www.humanservices.alberta.ca/documents/spf-summing-up-early-childhood-

    development.pdf

Alkin, M., & Christie, C. (2004). An evaluation theory tree. In M. Alkin, *Evaluation roots* (pp.

    13–65). Thousand Oaks, CA: SAGE. Retrieved from

    https://doi.org/10.4135/9781412984157.n2

Alkin, M. C. (2018). *Evaluation Essentials: From A to Z* (2nd ed.). New York, NY: Guilford
    Press.

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation
    influence, and factors affecting use. *American Journal of Evaluation*, *38*(3), 434–450.
    Retrieved from https://doi.org/10.1177/1098214017717015

American Educational Research Association. (2011). *Code of ethics, American Educational
    Research Association, approved by the AERA Council February 2011*. Retrieved from
    http://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics(1).pdf

American Educational Research Association, American Psychological Association, & National
    Council on Measurement in Education. (2014). *Standards for educational and
    psychological testing*. Washington, DC: American Educational Research Association.

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. *Journal of Evaluation*, *2*(27), 257–269. Retrieved from
    https://doi.org/10.1177/1098214006287989

Bakken, L. L., Núñez, J., & Couture, C. (2014). A course model for building evaluation capacity
    through a university–community partnership. *American Journal of Evaluation*, *35*(4),
    579–593. Retrieved from https://doi.org/10.1177/1098214014523671

Biott, C., & Cook, T. (2000). Local evaluation in a National Early Years Excellence Centres
    pilot programme: Integrating performance management and participatory evaluation.
    Evaluation, *6*(4), 399–413.

Blair, J., Czaja, R. F., & Blair, E. A. (2013). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Thousand Oaks, CA: SAGE.

Blumer, H. (1954). What is wrong with social theory? *American Sociological Review*, *19*(1), 3–10.

Botcheva, L., White, C. R., & Huffman, L. C. (2002). Learning culture and outcomes measurement practices in community agencies. *American Journal of Evaluation*, *23*(4), 421–434.

Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *Journal of Evaluation*, *34*(3), 299–319. Retrieved from https://doi.org/10.1177/1098214013477235

Bourgeois, I., Chouinard, J. A., & Cousins, J. B. (2008). Understanding organization capacity for evaluation: Synthesis and integration. *The Canadian Journal of Program Evaluation*, *23*(3), 225–237.

Boyle, R., & Lemaire, D. (1999). *Building effective evaluation capacity: Lessons from practice* (Vol. 8). New Brunswick, NJ: Transaction.

Brandon, P. R., & Fukunaga, L. L. (2014). The State of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, *35*(1), 26–44. Retrieved from https://doi.org/10.1177/1098214013503699

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bunge, M. (2012). *Scientific research II: The search for truth*. New York, NY: Springer Science & Business Media.

Burke, W. W. (2017). *Organization change: Theory and practice*. Thousand Oaks, CA: SAGE.

Burke, W. W., Lake, D. G., & Paine, J. W. (2008). *Organization change: A comprehensive reader*. San Francisco, CA: John Wiley & Sons.

Burke, W. W., & Litwin, G. H. (1992). A causal model of organizational performance and change. *Journal of Management*, *18*(3), 523–545.

Burns, G. N., Kotrba, L. M., & Denison, D. R. (2013). Leader-culture fit: Aligning leadership and corporate culture. In J. Passmore (Ed.), *The Wiley-Blackwell handbook of the psychology of leadership, change, and organizational development* (pp. 113–128). West Sussex, UK: John Wiley & Sons.

Byrne, B. M., & Crombie, G. (2003). Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*, *2*(3), 177–203.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada. (2014). *Tri-council policy statement: Ethical conduct for research involving humans* (No. RR4-2/2014E- PDF). Ottawa, Ontario: Secretariat on Responsible Conduct of Research. Retrieved from http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/

Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation*, *28*(1), 60–75. Retrieved from https://doi.org/10.1177/1098214006296245

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. Retrieved from https://doi.org/10.1177/1098214009352361

Chan, E. K. H. (2014). Standards and guidelines for validation practices: Development and evaluation of measurement instruments. In B. D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 9–24). New York, NY: Springer.

Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation* (pp. 33–55). Thousand Oaks, CA: SAGE.

Cheng, S.-H., & King, J. A. (2016). Exploring organizational evaluation capacity and evaluation capacity building: A Delphi study of Taiwanese elementary and junior high schools. *American Journal of Evaluation*, *38*(4), 521–539. Retrieved from https://doi.org/10.1177/1098214016672344

Chouinard, J. A. (2013). The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*, *34*(2), 237–253.

Christie, C. A., & Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation*, *34*(3), 131–135.

Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, *17*(1), 31–43. Retrieved from http://dx.doi.org/10.1037/a0026975

Clinton, J. (2014). The true impact of evaluation: Motivation for ECB. *American Journal of Evaluation*, *35*(1), 120–127. Retrieved from https://doi.org/10.1177/1098214013499602

Cohen, P., West, S. G., & Aiken, L. S. (2014). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Collie, R. J., & Zumbo, B. D. (2014). Validity evidence in the *Journal of Educational Psychology*: Documenting current practice and a comparison with earlier practice. In Bruno D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 113–135). New York, NY: Springer.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.

Cousins, J. B. (2004). Crossing the bridge: Toward understanding use through systematic inquiry. In M. C. Alkin (Ed.), *Evaluation roots: Tracing theorists' views and influences* (pp. 319–330). Thousand Oaks, CA: SAGE.

Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004). Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation*, *19*(2), 99–141.

Cousins, J. B., Elliott, C., Amo, C., Bourgeois, I., Chouinard, J., Goh, S. C., & Lahey, R. (2008). Organizational capacity to do and use evaluation: Results of a Pan-Canadian survey of evaluators. *The Canadian Journal of Program Evaluation*, *23*(3), 1–35.

Cousins, J. B., Goh, S. C., Elliott, C., Aubry, T., & Gilbert, N. (2014). Government and voluntary sector differences in organizational capacity to do and use evaluation. *Evaluation and Program Planning*, *44*, 1–13. Retrieved from https://doi.org/10.1016/j.evalprogplan.2013.12.001

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, *2014*(141), 7–23. Retrieved from https://doi.org/10.1002/ev.20076

Cousins, J. B., Bourgeois, I., & Associates. (2014). Multiple case study methods and findings. *New Directions for Evaluation*, *2014*(141), 25–99. Retrieved from https://doi.org/10.1002/ev.20077

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: SAGE.

Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: SAGE.

Creswell, J. W., & Plano-Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Los Angeles, CA: SAGE.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. Retrieved from https://doi.org/doi:10.1007/bf02310555

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Thousand Oaks, CA: SAGE.

Daigneault, P.-M., & Jacob, S. (2009). Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation*, *30*(3), 330–348. https://doi.org/10.1177/1098214009340580

Davies, P., Newcomer, K., & Soydan, H. (2006). Government as structural context for evaluation. In I. Shaw, J. Greene, & M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 163–183). Thousand Oaks, CA: SAGE.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. (L. Bickman & D. J. Rog, Eds.). Thousand Oaks, CA: Sage.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Hoboken, NJ: John Wiley & Sons.

Duignan, P. (2003). Mainstreaming evaluation or building evaluation capability? Three key elements. *New Directions for Evaluation*, *2003*(99), 7–21. Retrieved from https://doi.org/10.1002/ev.90

Early Child Development Mapping Project. (2014). *How are our young children doing? Final report of the Early Child Development Mapping Project*. Edmonton, Alberta: ECMap. Retrieved from http://www.ecmap.ca/images/results/ECMap_Final_Report_20141118.pdf

Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods*, *6*(4), 352.

Fabrigar, L. R. (2012). *Exploratory factor analysis*. Oxford, UK: Oxford University Press.

Ferns, C., & Friendly, M. (2014). *The state of early childhood education and care in Canada 2012*. Moving Childcare Forward Project. Retrieved from Movingchildcareforward.ca.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Thousand Oaks, CA: Sage.

Fierro, L. A. (2012). *Clarifying the connections: Evaluation capacity and intended outcomes* (Doctoral dissertation, The Claremont Graduate University). Available from ProQuest Dissertation & Theses Global database. (3545210)

Fierro, L. A., & Christie, C. A. (2017). Evaluator and program manager perceptions of

    evaluation capacity and evaluation practice. *American Journal of Evaluation*, *38*(3), 376–

    392. Retrieved from https://doi.org/10.1177/1098214016667581

Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action*

    approach. New York, NY: Psychology Press.

Fleming, M. A. (2011). Attitude, persuasion, and social influence: Applying social psychology to

    increase evaluation use. In M. M. Mark, S. I. Donaldson, & B. Campbell (Eds.), *Social*

    psychology and evaluation (pp. 212–242). New York, NY: Guilford Press.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of

    clinical assessment instruments. *Psychological Assessment*, *7*(3), 286–299. Retrieved

    from https://doi.org/10.1037/1040-3590.7.3.286

FluidSurvey. (2015). *Online Survey Software*. Ottawa. Retrieved from http://fluidsurveys.com

Gagnon, F., Aubry, T., Cousins, J. B., Goh, S. C., & Elliott, C. (2018). Validation of the

    evaluation capacity in organizations questionnaire. *Evaluation and Program Planning*,

    68, 166–175. Retrieved from https://doi.org/10.1016/j.evalprogplan.2018.01.002

Galport, N., & Azzam, T. (2017). Evaluator training needs and competencies: A gap analysis.

    American Journal of Evaluation, *38*(1), 80–100. Retrieved from

    https://doi.org/10.1177/1098214016643183

Gamst, G., Meyers, L. S., & Guarino, A. J. (2008). *Analysis of variance designs: A conceptual*

    and computational approach with SPSS and SAS. New York, NY: Cambridge University

    Press.

Ghoshal, S., & Bartlett, C. A. (1994). Linking organizational context and managerial action: The

    dimensions of quality management. *Strategic Management Journal*, *15*, 91–112.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review*, *17*(2), 183–211.

Goh, S. C., Cousins, J. B., & Elliott, C. (2006). Organizational learning capacity, evaluative inquiry and readiness for change in schools: Views and perceptions of educators. *Journal of Educational Change*, *7*(4), 289–318. Retrieved from https://doi.org/10.1007/s10833-005-5033-y

Goh, S. C., Quon, T. K., & Cousins, J. B. (2007). The organizational learning survey: A re-evaluation of unidimensionality. *Psychological Reports*, *101*(3), 707–721. Retrieved from https://doi.org/10.2466/pr0.101.3.707-721

Gokiert, R., Kingsley, B., Poth, C., Edwards, K., El Hassar, B., Tink, L., . . . Hopkins, S. (2017). Developing an evaluation capacity building network in the field of early childhood development. *Engaged Scholar Journal: Community-Engaged Research, Teaching, and Learning*, *3*(2), 59–80.

Gokiert, R. J., Noble, L., & Baugh Littlejohns, L. (2013). Enhancing early childhood (EC) professionals and Knowledge and practice of measurement. *NHSA Dialog*, *16*(3), 142–148. Retrieved from https://journals.uncc.edu/dialog/article/view/117

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: John Wiley & Sons.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., & Singer, E. (2004). *Survey methodology* (Vol. 1). Hoboken, NJ: Wiley.

Gunnell, K. E., Schellenberg, B. J., Wilson, P. M., Crocker, P. R., Mack, D. E., & Zumbo, B. D. (2014). A review of validity evidence presented in the *Journal of Sport and Exercise Psychology* (2002–2012): Misconceptions and recommendations for validation research.

In Bruno D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 137–156). New York, NY: Springer.

Harris, M. J. (2017). *Health program planning and evaluation*. Hoboken, NJ: John Wiley & Sons.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Hubley, A. M., Zhu, S. M., Sasaki, A., & Gadermann, A. M. (2014). Synthesis of validation practices in two assessment journals: *Psychological Assessment* and the *European Journal of Psychological Assessment*. In B. D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 193–213). New York, NY: Springer.

Hubley, A. M., & Zumbo, B. D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, *103*(2), 219–230. Retrieved from https://doi.org/10.1007/s11205-011-9843-4

Hubley, A. M., & Zumbo, B. D. (2013). Psychometric characteristics of assessment procedures: An overview. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 3–19). Washington, DC: American Psychological Association.

Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation*, *29*(3), 358–368. Retrieved from https://doi.org/10.1177/1098214008318798

Janzen, R., Ochocka, J., Turner, L., Cook, T., Franklin, M., & Deichert, D. (2017). Building a community-based culture of evaluation. *Evaluation and Program Planning*, *65*, 163–170.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, *1*(2), 112–133. Retrieved from https://doi.org/10.1177/1558689806298224

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, *38*(4), 319–342. Retrieved from https://doi.org/10.1111/j.1745-3984.2001.tb01130.x

Kane, M. T. (2013a). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. Retrieved from https://doi.org/10.1111/jedm.12000

Kane, M. T. (2013b). Validation as a pragmatic, scientific activity. *Journal of Educational Measurement*, *50*(1), 115–122. Retrieved from https://doi.org/10.1111/jedm.12007

Kemper, E. A., Stringfield, S., & Teddlie, C. (2003). Mixed methods sampling strategies in social science research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 273–296). Thousand Oaks, CA: SAGE.

King, J. A. (2007). Developing evaluation capacity through process use. *New Directions for Evaluation, 2007*(116), 45–59. Retrieved from https://doi.org/10.1002/ev.242

King, J. A. (in press). Putting evaluation capacity building in context: Reflections on the Ontario Brain Institute's Evaluation Support Program. *Evaluation and Program Planning*. Retrieved from https://doi.org/10.1016/j.evalprogplan.2017.05.013

King, J. A., & Sevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. Los Angeles, CA: SAGE.

King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter*, *35*(3), 10–16.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). New York, NY: Guilford Press.

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher & C. Schatsschneider (Eds.), *Applied quantitative analysis in the social sciences* (pp. 171–207). New York, NY: Routledge.

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. Retrieved from https://doi.org/10.1177/1098214013499965

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. Retrieved from https://doi.org/10.1177/1098214011434608

Lamm, A. J., & Israel, G. D. (2011). An organizational approach to understanding evaluation in extension. *Journal of Agricultural Education*, *52*(4), 136–149. Retrieved from https://doi.org/10.5032/jae.2011.04136

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, *73*(5), 467–482. https://doi.org/10.1016/j.techfore.2005.09.002

Langford, R., & Richardson, B. (2018). Early childhood education and care in Canada: Consistently inconsistent childcare policy. In S. Garvis, S. Phillipson, & H. Harju-Luukkainen (Eds.), *International perspectives on early childhood education and care* (pp. 38–51). New York, NY: Routledge.

Latcheva, R. (2011). Cognitive interviewing and factor-analytic techniques: A mixed method

    approach to validity of survey items measuring national identity. *Quality & Quantity*,

    45(6), 1175–1199. Retrieved from https://doi.org/10.1007/s11135-009-9285-0

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum

    likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3),

    936–949. Retrieved from https://doi.org/10.3758/s13428-015-0619-7

Liket, K. C., Rey-Garcia, M., & Maas, K. E. H. (2014). Why aren't evaluations working and

    what to do about it: A framework for negotiating meaningful evaluation in nonprofits.

    American Journal of Evaluation, *35*(2), 171–188. Retrieved from

    https://doi.org/10.1177/1098214013517736

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to

    parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, *9*(2),

    151–173.

Luyt, R. (2012). A framework for mixing methods in quantitative measurement development,

    validation, and revision: A case study. *Journal of Mixed Methods Research*, *6*(4), 294–

    316. Retrieved from https://doi.org/10.1177/1558689811427912

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in

    psychological research. *Annual Review of Psychology*, *51*(1), 201–226. Retrieved from

    https://doi.org/10.1146/annurev.psych.51.1.201

Mark, M. M., Greene, J. C., & Shaw, I. F. (2006). Introduction: The evaluation of policies,

    programs and practices. In *The Sage handbook of evaluation* (pp. 1-30). Thousand Oaks,

    CA: SAGE.

Masten, A. S., & Gewirtz, A. H. (2006). Resilience in development: The importance of early childhood. In R. E. Tremblay, R. G. Barr, & R. D. Peters (Eds.), *Encyclopedia on early childhood development* (pp. 1–6). Montreal, Quebec: Centre of Excellence for Early Childhood Development. Retrieved from http://www.child-encyclopedia.com/Pages/PDF/Masten-GewirtzANGxp.pdf

McCoy, A., Rose, D., & Connolly, M. (2013). Developing evaluation cultures in human service organisations. *Evaluation Journal of Australasia*, *13*(1), 15–20. Retrieved from https://doi.org/10.1177/1035719X1301300103

Meadows, P. D. (1998). *An evaluation of community capacity building: Assessing methods and processes within a community education program* (Doctoral dissertation, Wayne State University). Available from ProQuest Dissertation & Theses Global database. (9915706)

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–104). New York, NY: Macmillan.

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*, *2002*(93), 27–46. Retrieved from https://doi.org/10.1002/ev.40

Moss, P. A. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, *17*(2), 6–12.

Muijs, D., Aubrey, C., Harris, A., & Briggs, M. (2004). How do they manage? A review of the research on leadership in early childhood. *Journal of Early Childhood Research*, *2*(2), 157–169. Retrieved from https://doi.org/10.1177/1476718X04042974

Nakaima, A., & Sridharan, S. (in press). Reflections on experiential learning in evaluation capacity building with a community organization, Dancing With Parkinson's. *Evaluation*

and Program Planning. Retrieved from

https://doi.org/10.1016/j.evalprogplan.2017.05.002

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. Retrieved from https://doi.org/10.1177/1098214010396075

Patton, M. Q. (2014). What brain sciences reveal about integrating theory and practice. *American Journal of Evaluation*, *35*(2), 237–244. https://doi.org/10.1177/1098214013503700

Patton, M. Q. (2011a). *Essentials of utilization-focused evaluation*. Thousand Oaks, CA: SAGE.

Patton, M. Q. (2011b). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. Guilford Press.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.

Preskill, H. (2013). Now for the hard stuff: Next steps in ECB research and practice. *American Journal of Evaluation*, *35*(1), 116–119. Retrieved from

https://doi.org/10.1177/1098214013499439

Preskill, H., & Boyle, S. (2008a). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, *29*(4), 443–459. Retrieved from

https://doi.org/10.1177/1098214008324182

Preskill, H., & Boyle, S. (2008b). Insights into evaluation capacity building: Motivations, strategies, outcomes, and lessons learned. *Journal of Program Evaluation*, *23*(3), 147–174.

Preskill, H., & Torres, R. T. (2001). Organizational readiness for learning and evaluation. In O.

    A. Aliaga (Ed.), *The 2001 AHRD conference proceedings* (Vol. 1, pp. 704–719). Tulsa,

    OK: Academy of Human Resource Development. Retrieved from

    http://sites.google.com/site/reyty1/2001Proceedings.pdf#page=705

Punch, K. F. (2013). *Introduction to social research: Quantitative and qualitative approaches*.

    Thousand Oaks, CA: SAGE.

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY:

    Routledge.

Runnels, V., Andrew, C., & Rae, J. (2017). Building evaluation culture and capacity in a

    community-level program: Lessons learned from evaluating youth futures. *Canadian*

    Journal of Program Evaluation, *32*(1), 122–130. Retrieved from

    https://doi.org/10.0.12.66/cjpe.30975

Sanders, J. R. (2003). Mainstreaming evaluation. *New Directions for Evaluation*, *2003*(99), 3–6.

    Retrieved from https://doi.org/10.1002/ev.89

Sauer, P. L., & Dick, A. (1993). Using moderator variables in structural equation models. In L.

    McAlister and M. L. Rothschild (Eds.), in *NA - Advances in Consumer Research,* (Vol.

    20, pp. *636-640.*636–640). Provo, UT: Association for Consumer Research. Retrieved

    from http://acrwebsite.org/volumes/7532/volumes/v20/NA-20

Schumacker, R. E., & Lomax, R. G. (2012). *A beginner's guide to structural equation modeling*

    (3rd ed.). New York, NY: Routledge.

SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in

    evaluation: An overview. *New Directions for Evaluation*, *2004*(102), 5–19. Retrieved

    from https://doi.org/10.1002/ev.112

Shaw, I. F., Greene, J. C., & Mark, M. M. (Eds.). (2006). *The Sage handbook of evaluation*. Thousand Oaks, CA: SAGE.

Shepard, L. A. (2016). Evaluating test validity: Reprise and progress. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 268–280. Retrieved from https://doi.org/10.1080/0969594X.2016.1141168

Shonkoff, J. P. (2017). Rethinking the definition of evidence-based interventions to promote early childhood development. *Pediatrics*, *140*(6), 31–36. Retrieved from https://doi.org/10.1542/peds.2017-3136

Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). Introduction. In *From neurons to neighborhoods: The science of early childhood development* (pp. 19–38). Washington, DC: National Academies Press.

Shear, B., & Zumbo, B. D. (2014). What counts as evidence: A review of validity studies in educational and psychological measurement. In Bruno D. Zumbo & E. K. H. Chan (Eds.), *Validity and validation in social, behavioral, and health sciences* (pp. 91–112). New York, NY: Springer.

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.

Sireci, S. G., & Sukin, T. (2013). Test validity. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds), *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology* (pp. 61–84). Washington, DC: American Psychological Association. Retrieved from http://dx.doi.org/10.1037/14047-004

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–898. Retrieved from https://doi.org/10.1016/j.paid.2006.09.017

Stockdill, H. S., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, *2002*(93), 7–26. Retrieved from https://doi.org/10.1002/ev.39

Suarez-Balcazar, Y., Taylor-Ritzler, T., Garcia-Iriarte, E., Keys, C. B., Kinney, L., Rush-Ross, H., . . . Curtin, G. (2010). Evaluation capacity building: A cultural and contextual framework. In F. Balcazar, Y. Suarez-Balcazar, T. Taylor-Ritzler, C. B. Keys (Eds.), *Race, Culture, and Disability: Rehabilitation Science and Practice* (pp. 307–324). Boston, MA: Jones & Bartlett.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2010). *Early childhood matters: Evidence from the effective pre-school and primary education project*. New York, NY: Routledge.

Tabachnick, B. G., & Fidell, L. S. (2006). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: SAGE.

Taut, S. (2007). Defining evaluation capacity building: Utility considerations. *American Journal of Evaluation*, *28*(1), 120. Retrieved from https://doi.org/10.1177/1098214006298062

Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.),

Program evaluation in a complex organizational system: Lessons from Cooperative Extension [Special issue]. *New Directions for Evaluation, 120,* 55–69.

Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D. B., & Balcazar, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation*, *34*(2), 190–206. Retrieved from https://doi.org/10.1177/1098214012471421

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Thurstone, L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529–554.

Touré-Tillery, M., & Fishbach, A. (2014). How to measure motivation: A guide for the experimental social psychologist. *Social and Personality Psychology Compass*, *8*(7), 328–341.

Trawick-Smith, J. (2013). *Early childhood development: A multicultural perspective* (6th ed.). New York, NY: Pearson.

Treiblmaier, H., & Filzmoser, P. (2010). Exploratory factor analysis revisited: How robust methods support the detection of hidden multivariate data structures in IS research. *Information & Management*, *47*(4), 197–207. Retrieved from https://doi.org/10.1016/j.im.2010.02.002

Visser, P. S., Krosnick, J. A., & Lavarakas, P. J. (2000). Survey research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 223–252). New York, NY: Cambridge University Press.

Volkov, B. B., & King, J. A. (2007). *A checklist for building organizational evaluation capacity*. Retrieved from https://goo.gl/Tvf1Nf

Walker-Egea, C. F. (2014). *Design and validation of an evaluation checklist for organizational readiness for evaluation capacity development* (Doctoral dissertation, University of South Florida, Ann Arbor). Available from ProQuest Dissertation & Theses Global. (5399)

Wandersman, A. (2014). Moving forward with the science and practice of evaluation capacity building (ECB): The why, how, what, and outcomes of ECB. *Journal of Evaluation*, *35*(1), 87–89. Retrieved from https://doi.org/10.1177/1098214013503895

Weitzman, B. C., & Silver, D. (2013). Good evaluation measures: More than their psychometric properties. *American Journal of Evaluation*, *34*(1), 115–119. Retrieved from https://doi.org/10.1177/1098214012461628

White, L. A., & Prentice, S. (2016). Early childhood education and care reform in Canadian provinces: Understanding the role of experts and evidence in policy change. *Canadian Public Administration*, *59*(1), 26–44.

Wiersma, W., & Jurs, S. G. (2009). *Research methods in education: An introduction*. New York, NY: Pearson/Allyn and Bacon.

Wilcox, Y., & King, J. A. (2013). A professional grounding and history of the development and formal use of evaluator competencies. *The Canadian Journal of Program Evaluation*, *28*(3), 1–28.

Williams, B., & Imam, I. (Eds.). (2007). Systems thinking for evaluation. In *Systems concepts in evaluation: An expert anthology*. Point Reyes, CA: EdgePress of Inverness.

Wolf, C., Joye, D., Smith, T., & Fu, Y. (2016). *The SAGE handbook of survey methodology*. London, UK: SAGE. Retrieved from http://sk.sagepub.com/Reference/the-sage-handbook-of-survey-methodology

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. Thousand Oaks, CA: SAGE.

Zumbo, B. D. (2009). Validity as a contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 65–82). Charlotte, NC: Information Age.

Zumbo, B. D., & Chan, E. K. H. (2014). *Validity and validation in social, behavioral, and health sciences*. New York, NY: Springer.

Appendix A. Operationalizing Evaluation Capacity: Protocol for Session with the Core

Research Team

April 13, 2015
(Allocated time: 55 min)

1. **Preamble (10 min)**
   - The purpose of the working session

   - Developing evaluation capacity (EC) framework that is contextually relevant, based on literature and community

   - Current EC and EC building literature
     o Retroactive
     o Only one valid instrument
     o EC frameworks have not been tested (only one)

   - The value of our approach
     o Develop context-appropriate EC
     o As opposed to retroactive, empirically develop EC and study it
     o Measure the components through the EC instrument

   - This session is about the EC components and their definitions
     o The following components are based on the PMT working session
     o Synthesis of the existing frameworks
     o Discuss four example frameworks

   - We should remember that the focus is evaluation capacity and not evaluation

2. **Evaluation capacity components review (15 min)**
   - Introduction (2 min)
   - Review components individually (5 min)
   - Review components as a group (8 min)

Guiding questions:
   - What components/factors are necessary to develop and implement evaluation capacity?
   - Are the current wordings of the components accurate?
   - Any other suggestions?

3. **Definitions of evaluation capacity components (25 min)**
   - Introduction (2 min)
   - Review definitions individually (8 min)
   - Review definitions as a group (15 min)

Guiding questions:

- Do the current definitions accurately describe the components? Is there anything missing in the definition?
- Any other suggestions?

4. **Conclusion and next steps (Less than 5 min)**
   - Overall purpose of this research and its potential
   - Items and scale construction
   - Visual representation of the framework (representing the relationships)

Components and their Definitions Based on the Literature

The purpose of the framework and instrument: Determine the current evaluation capacity level, needs, and priorities of organizations working in the early childhood field and their employees.

To what extent do you agree with the inclusion of each of the framework components and their definitions? Please rate each component/definition using the following scale:

D: Disagree   **N**: Neutral (neither agree or disagree)   **A**: Agree

Framework Component	Framework Rating	Operational Definitions and Example Items	Definition Rating	Suggestions
Evaluation capacity (EC) building Strategies to develop capacity	*D N A*	*It includes individual and organizational strategies to develop evaluation capacity. These strategies may include tools and resources made available by the organizations and types of activities considered.* ***Example items*** *To what extent do you agree with the following statements:* *1. My organization has procedures in place to facilitate professional development.*	*D N A*	
Mediators	*D N A*	These refer to features that may influence EC outcomes such as resource availability, external requirements (e.g., funders, stakeholders), timeliness of evaluation (e.g., what prompted a need for evaluation, timing of the evaluation, and frequency). ***Example items*** *To what extent do you agree with the following statements:* *1. I engage in evaluation activities that are consistent with funders' expectations.* *2. I perform evaluation activities frequently as part of the program implementation.*	*D N A*	

	D N A		D N A	
Individual capacity	*D* *N* *A*	It refers to individual's evaluation-related competencies (e.g., evaluation technical knowledge and skills, soft skills), individual readiness (e.g., attitudes and motivations toward evaluation), and individual contextual awareness (e.g., understanding of organizational needs). ***Example items*** *To what extent do you agree with the following statements:* 1. *I am motivated to learn about evaluation.* 2. *I support other colleagues to evaluate their program.* 3. *I know how to develop an evaluation plan.* 4. *I know how to define measurable program outcomes.* 5. *I developed an evaluation plan in the past.*	*D* *N* *A*	
Organizational learning culture	*D* *N* *A*	It refers to the learning climate of the organization and the role of evaluation in decision-making. Key elements of this component are effectiveness of leadership, embracing innovation and experimentation, alignment between organizational mission, organizational values, and norms that are conducive to EC and performing evaluations. ***Example items*** *To what extent do you agree with the following statements:* 1. *The leadership of my organization encourages me to be innovative in my work.* 2. *Executive leaders in my organization support and value program evaluation and evaluative thinking.* 3. *Evaluation information in my organization is shared openly with all staff.* 4. *My organization regularly (e.g., annually) assesses compatibility between existing programs and the organization's mission.*	*D* *N* *A*	
Evaluative thinking	*D* *N* *A*	It refers to a "type of reflective practice that incorporates use of systematically collected data to inform organizational decisions and other actions" (Baker & Bruner, 2006, p. 34). This could include increasing ownership of evaluation results and commitment to using its results, providing incentives, rewards, and recognitions related to evaluation, and integrating evaluation into work practices, policies, and decision-making. ***Example items***	*D* *N* *A*	

			To what extent do you agree with the following statements: 1. *I am able to integrate evaluation activities into my daily work practices.* 2. *I regularly use the evaluation results to make my work-related decisions.*			
Organizational capacity to do evaluation	*D*	*N*	*A*	It refers to the elements necessary to conduct evaluation (e.g., knowledge and skills, resources). ***Example items:*** *To what extent do you agree with the following statements:* 1. *I have time to conduct evaluation activities (e.g., identifying or developing a survey, collecting information from participants).* 2. *I have access to technology to compile information into computerized records.* 3. *I have access to learning opportunities to develop evaluation skills I need (e.g., trainings, workshops, online resources).*	*D*	*N*	*A*	
Capacity to use evaluation	*D*	*N*	*A*	It refers to the elements necessary to use evaluation results in, for example, program planning and decision-making. ***Example items*** *To what extent do you agree with the following statements:* 1. *I use evaluation results to report to a funder's requirements.* 2. *I use evaluation results to improve the program(s).* 3. *I use evaluation results to make informed decisions.*	*D*	*N*	*A*	
Organizational context	*D*	*N*	*A*	It includes "organization's history, environment, type, relationship with the community, and ways in which it operates" (Suarez-Balcazar et al., 2010, p. 314). ***Example items*** *To what extent do you agree with the following statements:* 1. *My program gathers information from diverse stakeholders to understand how well the program is doing.* 2. *My program has adequate records of past evaluation efforts.*	*D*	*N*	*A*	

Appendix B. Information Letters and Consent Forms

This appendix includes four parts: 1) Information letter for expert review, 2) Recruitment to Attend the Dialogue Forums and Complete the Evaluation Capacity Instrument, 3) Invitation Letter to Complete Evaluation Capacity Instrument (second administration), 4) Online Information and Consent Form.

1. Information letter for expert review

Study Title: Evaluation Capacity Network: Aligning Evaluative Thinking and Practice among Early Childhood Stakeholders

Principal Investigator: Dr. Rebecca Gokiert (rgokiert@ualberta.ca or 780-492-6297)
Community-University Partnership for the Study of Children, Youth, and Families, 2-281 Enterprise Square,
10230 Jasper Ave., University of Alberta, Edmonton, AB, T5J 4P6.

- Your view is important to us! You are invited to participate in this study because your work is related to evaluation within early childhood field.
- This research is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

Purpose
- discussions among early childhood stakeholders (government agencies, community organizations, funders, and academics) have shown that there is a lack of central coordination of evaluation resources and expertise accessible to early childhood stakeholders to ensure high quality research, training, and practice in early childhood evaluation.
- the current project aims to engage early childhood stakeholders across disciplines, sectors, and provinces in a participatory research process in which knowledge is co-created and mobilized to align evaluative thinking and enhance evaluation capacity in the field of early childhood development. this will be achieved through the three main objectives:
  1. to conduct an intersectoral needs assessment using community forums with the aim of identifying common evaluation knowledge and capacity gaps
  2. to develop and deliver educational resources and training opportunities that address these gaps and then to evaluate and refine the resources and training
  3. to study the process of establishing an evaluation capacity network partnership that supports ongoing dialogue of early childhood stakeholders and knowledge mobilization of community-engaged evaluative practices across sectors.

Research activities
- *Provincial community forums:* The community forums will be from 8:30 am to 4:30 pm. During the forums, there will be a number of one-hour focus groups on common evaluation knowledge and capacity gaps. At the end of the forums, participants will fill out a post-forum participant survey on the effectiveness of the community forums, which will take about 15 minutes.
- *Feedback on the summary report of the community forums:* Participants of the forums are invited to provide feedback on the summary report that will be shared on an online platform or at various knowledge-sharing events. Each knowledge-sharing event will be 3 hours, including presentation of the report, Q & A, and a written survey. The survey will take no more than 30 minutes to complete.
- *Mid-point partner interviews and surveys:* In addition to participation in community forums and feedback on the summary report, members of ECN governing bodies (Steering Committee, Core Research Team, Project Management Team) will also participate in mid-point partner interviews and surveys to assess best practices in establishing and sustaining an intersectoral community–university

partnership. Interviews will be about one hour long, and surveys will take no more than 15 minutes to complete.

Benefits and Risks
Participants will:
- expand their understanding of the evaluation challenges and issues currently being experienced across sectors
- translate and utilize the outcomes of the needs assessment within their own sectors
- foster intersectoral relationships that will build capacity initiatives
- stimulate positive change in evaluation in the field of early childhood
- not experience any risk by taking part in this research.

Voluntary participation
- You do not have to participate in this study. Your participation is up to you. You may choose to withdraw from the research at any time.
- You may contact the Principal Investigator, Dr. Rebecca Gokiert, to ask to have your survey and interview data removed within 2 weeks following your submission. However, focus group data will be not able to be removed because the focus group discussions will be audio recorded and it will be too difficult to determine your individual contribution after the fact.

Confidentiality and Anonymity
- Research results from this study may be used for presentations and publications, teaching purposes, and secondary data analyses that use anonymous data.
- All information gathered will be confidential. The researchers and all others involved in the research will comply with the University of Alberta Standards for the Protection of Human Research Participants. They will sign a confidentiality agreement.
- Quotes from focus group discussions or feedback on the summary report will be identified by sector (i.e., community, university, government, funder) in any documents resulting from this research. Consent information will be kept in a separate file from your data.
- Because of the nature of participatory research activities, participants usually meet in group settings (e.g., focus groups during provincial forums, knowledge sharing events), where total anonymity cannot be guaranteed. To assure confidentiality, participants will sign a confidentiality agreement so that no information related to any participant will be brought outside of the research activities.
- Your data will be kept in a secure digital file or cabinet. After 5 years of when the project is completed, we will destroy your data.

If you have any questions or concerns, please contact Dr. Rebecca Gokiert at rgokiert@ualberta.ca or 780-492-6297.
The plan for this study has been reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta. For questions regarding participant rights and ethical conduct of research, please contact the Research Ethics Office at (780) 492-2615.
Consent Statement

I have read this form and the research study has been explained to me. I have been given the opportunity to ask questions and my questions have been answered. If I have additional questions, I have been told whom to contact. I agree to participate in the research study described above and will receive a copy of this consent form. I will receive a copy of this consent form after I sign it.

_____   _____
Participant's Name (Printed) and Signature   Date

_____   _____
Name (Printed) and Signature of Person Obtaining Consent   Date

2. Recruitment to Attend the Dialogue Forums and Complete the Evaluation Capacity Instrument

Study Title: Evaluation Capacity Network: Aligning Evaluative Thinking and Practice among Early Childhood Stakeholders

Principal Investigator (Researcher): Dr. Rebecca Gokiert, Community-University Partnership for the Study of Children, Youth, and Families (CUP)

Purpose: Through discussions with our community partners, including government agencies, community organizations, funders, and academics in the field of early childhood field in Alberta, we found that there is a lack of central coordination of evaluation resources and expertise accessible to early childhood stakeholders to ensure high quality research, training, and practice in early childhood evaluation. the community-university partnership at the University of Alberta invites you to take part in a discussion about evaluation capacity building. the purpose of the discussion is to help us understand the current need for evaluation capacity building and identify best ways to address it.

What and who is involved? Individuals involved with early childhood oganizations and groups who are involved in any aspect of evaluation (e.g., creating or approving budgets for evaluation, designing evaluation plans, collecting and analyzing data, creatinh evaluation reports, and disseminating findings; and/or managing or overseeing evaluations performed externally). You are being invited to attend a final event for this study where our findings will be shared back with all of our partners and those who participated throughout the study. Your input during the knowledge translation event will be used to inform the research and create capacity-building resources. At the end of the event you will be asked to complete an evaluation of the event. Completion of the evaluation is completely voluntary.

How will we protect your privacy? The information that you provide will be kept private. No names will be attached to the information or in any reports from the study. Since some of the discussions are happening in a group setting, confidentiality is a shared responsibility and is requested from all participants, but it cannot be guaranteed. The data from the discussions will be kept in a locked filing cabinet at the University of Alberta and will only be available to the research team. It is our intent to summarize the results from these discussions and share it with people in the early childhood community. We may publish the overall results from this study in scholarly journals and present results at conferences; however, individual participant comments will not be identifiable because all discussion data will be combined. We will keep the data for a minimum of 5 years and then the data will be destroyed in a way that ensures privacy and confidentiality. If the data are used for other studies, such as for a students' thesis work, ethics approval will be obtained.

Contact Names and Telephone Numbers:
Thank you very much for considering this request. If you have any questions or would like more information about the study, please contact me at rgokiert@ualberta.ca or 780-492-6297.The plan for this study has been reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta. For questions regarding participant rights and ethical conduct of research, please contact the Research Ethics Office at (780) 492-2615.

Sincerely,

Rebecca Gokiert, PhD, R.Psych
Associate Professor, Faculty of Extension
Assistant Director, Community-University Partnership (CUP)

3. Invitation Letter to Complete Evaluation Capacity Instrument (second administration)

Dear [insert 'Contact Name']:

Purpose: I am writing on behalf of the Community-University Partnership at the University of Alberta to request 20 minutes of your time to respond to complete an instrument about evaluation capacity building. You were sent this invitation letter because you either work at an early childhood development (early childhood) organization or are connected to early childhood organizations through your work. please send this invitation letter to employees or early childhood organizations responsible for any evaluation-related activities (e.g., overseeing evaluations, using evaluation information to inform practice, collecting evaluation data). If you have received this invitation in error and know of an appropriate respondent, please forward this email to that person.

Study Title: Evaluation Capacity Network: Aligning Evaluative Thinking and Practice among Early Childhood Stakeholders

Principal Investigator (Researcher): Dr. Rebecca Gokiert, Community-University Partnership for the Study of Children, Youth, and Families (CUP)

Background: Through discussions with our community partners, including government agencies, community organizations, funders, and academics in the field of early childhood field in alberta, we found that there is a lack of central coordination of evaluation resources and expertise accessible to early childhood stakeholders to ensure high-quality research, training, and practice evaluation within early childhood field. The survey will help us understand the current need for evaluation capacity building and identify best ways to address it.

Who should fill out the instrument: Employees in your organization who are involved in any aspect of completing evaluation, for example, those who create or approve budgets for evaluation, design evaluation plans, collect and analyze data, create evaluation reports, and share results, and/or manage or oversee evaluations performed externally.

Protecting your Privacy: Your participation is completely voluntary, and all of your responses will be kept confidential. Moreover, the results of all respondents will be aggregated. Therefore, no one will link you to your responses. Text comments will be reported verbatim, so please do not provide identifying information in your text comments.

EC instrument link: please click here to be directed to the online instrument.

Thank you in advance for your participation in this important project. If you have any questions please contact me at rgokiert@ualberta.ca or 780-492-6297.

Sincerely,

Rebecca Gokiert, PhD, R.Psych
Associate Professor, Faculty of Extension, University of Alberta

4. Online Information and Consent Form

You are being invited to participate in a research study titled ***"Evaluation Capacity Network: Aligning Evaluative Thinking and Practice among Early Childhood Stakeholders"***. The principal researcher is **Dr. Rebecca Gokiert,** Community-University Partnership for the Study of Children, Youth, and Families (CUP), University of Alberta. You have been selected to participate because your work is related to early childhood field and you are responsible for doing program evaluation and/or oversee evaluation performed externally.

The purpose of this project is to help us understand the current need for evaluation capacity building and identify best ways to address this need. If you agree to take part in this study, you will be asked to complete an online instrument.

You may not directly benefit from this research. however, we hope that by completing the instrument you may help the early childhood partners and network build evaluation capacity that meets the needs of early childhood organizations in alberta. we also hope that this study stimulates positive change in evaluation in the field of early childhood.

We believe there are no known risks associated with this research study. Your participation in the survey is completely voluntary, all of your responses will be kept confidential, and you can withdraw at any time. The results of all respondents will be aggregated when presented or published. Therefore, no one will link you to your responses. Text comments will be reported verbatim, so please do not provide identifying information in your text comments.

If you have any questions or concerns, please contact **Dr. Rebecca Gokiert** at rgokiert@ualberta.ca or 780-492-6297. The plan for this study has been reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta. For questions regarding participant rights and ethical conduct of research, please contact the Research Ethics Office at (780) 492-2615.

By clicking "I agree" below you are indicating that you are at least 18 years old, have read and understood this consent form, and agree to participate in this research study. Please print a copy of this page for your records.

| I agree | I do not agree |

Appendix C. Expert Group Review Protocol

Step 1- Brief Individual Review

- Please read through the Evaluation Capacity (EC) instrument, including the directions, items, and scales.
- Given your expertise (e.g., evaluation, measurement, community context), as you review the EC instrument, please insert the following at the beginning of items:

  o "unclear"
  o "has measurement issues"
  o "inappropriate to context"

Please note the questions in the EC instrument are organized into 3 major sections:
1. Section A – Individual Evaluation Capacity
2. Section B – Organizational Evaluation Capacity
3. Section C – Information About You and Your Organization

Step 2: Group Discussion

Through the use of a talking circle and enlarged EC instrument, for each section we will review: (1) instruction, (2) individual items, to discuss the following questions:

1. Is the wording of the instruction clear?

2. Are the sections appropriate from a cultural or contextual perspective? Why or why not?

3. Are there any specific questions that stood out to you as a problem? If yes, why do you feel they are a problem?

4. Are there any aspects of evaluation capacity that are not represented by specific questions? If yes, provide suggestions.

Appendix D. Evaluation Capacity Instrument (Final Version)

Evaluation Capacity Instrument

Evaluation, for the purpose of this survey, is defined as a process of systematically gathering information necessary to enhance programs, projects, or policies, and determine the extent to which program outcomes are being met. This survey aims to identify both **individual** and **organizational** elements believed to be critical for evaluation capacity building.

A.   Individual capacity

The following statements are intended to gather information on some key elements (e.g., knowledge, resources) that are important in understanding and thereby building evaluation capacity. Based on your understanding, **please indicate the extent to which you agree or disagree with each statement**.

	Strongly disagree	Disagree	Agree	Strongly agree	NA
Ind_1. I have an understanding of evaluation.	☐	☐	☐	☐	☐
Ind_2. I have a working knowledge of conducting an evaluation.	☐	☐	☐	☐	☐
Ind_3. I have a clear sense about the benefits of evaluation.	☐	☐	☐	☐	☐
Ind_4. I think evaluation is important for making improvements in my organization.	☐	☐	☐	☐	☐
Ind_5. I think evaluation can be useful in determining the effectiveness of my organization.	☐	☐	☐	☐	☐
Ind_6. I think evaluation findings can be beneficial to my sponsor(s)/funder(s).	☐	☐	☐	☐	☐
Ind_7. I think evaluation can improve transparency in an organization.	☐	☐	☐	☐	☐
Ind_8. I would be concerned if an organization doesn't evaluate its activities.	☐	☐	☐	☐	☐
Ind_9. I have an ethical responsibility to participate in evaluation as needed.	☐	☐	☐	☐	☐
Ind_10. I think it is important that staff members get involved in evaluation.	☐	☐	☐	☐	☐
Ind_11. I think evaluation creates unnecessary additional workload.	☐	☐	☐	☐	☐
Ind_12. I have the necessary resources (e.g., time, money) to engage in evaluation.	☐	☐	☐	☐	☐
Ind_13. I often seek out professional development opportunities related to evaluation.	☐	☐	☐	☐	☐
Ind_14. I think time spent on evaluation is justified.	☐	☐	☐	☐	☐
Ind_15. I think internal allocation of funding for an evaluation is justified.	☐	☐	☐	☐	☐
Ind_16. I think credentialed training (e.g., course, certificate, diploma) in evaluation is important.	☐	☐	☐	☐	☐
Ind_17. I think non-credentialed training (e.g., webinar, coaching) in evaluation is important.	☐	☐	☐	☐	☐
Ind_18. I think external funding for an evaluation is justified.	☐	☐	☐	☐	☐
Ind_19. I know what skills to look for in an external evaluator.	☐	☐	☐	☐	☐
Ind_20. I am open to learning the necessary skills myself for conducting an evaluation.	☐	☐	☐	☐	☐

Ind_21. I am open to staff being provided with the opportunities to learn the skills necessary in evaluation.	☐	☐	☐	☐	☐
Ind_22. I have the skills to oversee an external evaluator.	☐	☐	☐	☐	☐
Ind_23. I am open to adopting new ideas in my day-to-day activities based on evaluation findings.	☐	☐	☐	☐	☐
Ind_24. I know how to use evaluation findings in decision-making.	☐	☐	☐	☐	☐
Ind_25. I know how to make organizational level changes based on evaluation findings.	☐	☐	☐	☐	☐
Ind_26. I am open to adopting new practices or processes based on evaluation findings.	☐	☐	☐	☐	☐

The following items are intended to help us better understand your current evaluation skills and how these skills are important to your work. **For each item, please indicate your personal level of skill in conducting each evaluation activity and how important you believe each activity is to your work.**

	Skill Level			Importance		
	Not at All skilled	Somewhat skilled	Very skilled	Not at all important	Somewhat important	Very important
1a. Using evaluation terminology (e.g., logic model, stakeholder map)	☐	☐	☐	☐	☐	☐
2a. Involving stakeholders in setting evaluation goals	☐	☐	☐	☐	☐	☐
3a. Formulating an evaluation plan	☐	☐	☐	☐	☐	☐
4a. Developing evaluation questions	☐	☐	☐	☐	☐	☐
5a. Choosing the right data collection method(s) (e.g., interview, survey) for an evaluation	☐	☐	☐	☐	☐	☐
6a. Using qualitative data collection methods (e.g., interview)	☐	☐	☐	☐	☐	☐
7a. Using quantitative data collection methods (e.g., survey)	☐	☐	☐	☐	☐	☐
8a. Analyzing evaluation data quantitatively	☐	☐	☐	☐	☐	☐
9a. Analyzing evaluation data qualitatively	☐	☐	☐	☐	☐	☐
10a. Interpreting quantitative data to answer evaluation questions	☐	☐	☐	☐	☐	☐
11a. Interpreting qualitative data to answer evaluation questions	☐	☐	☐	☐	☐	☐
12a. Developing recommendations based on evaluation findings	☐	☐	☐	☐	☐	☐
13a. Writing an evaluation report	☐	☐	☐	☐	☐	☐

	Strongly disagree	Disagree	Agree	Strongly agree	Don't Know
14a. Presenting evaluation findings	☐	☐	☐	☐	☐
15a. Using evaluation findings to improve my work	☐	☐	☐	☐	☐
16a. Using evaluation results to make funding decisions	☐	☐	☐	☐	☐

B. Organizational capacity

The following statements are intended to gather information on some key aspects (e.g., resources) that are important for understanding how to build evaluation capacity at the organizational level. **Based upon your own experience at work, please choose the best option for each statement.**

	Strongly disagree	Disagree	Agree	Strongly agree	Don't Know
My organization…					
Org_1. Encourages staff to express their opinions	☐	☐	☐	☐	☐
Org_2. Involves staff when making long-term plans	☐	☐	☐	☐	☐
Org_3. Gives staff the opportunity to reflect on organizational goals	☐	☐	☐	☐	☐
Org_4. Provides professional development opportunities	☐	☐	☐	☐	☐
Org_5. Provides opportunities for learning evaluation skills	☐	☐	☐	☐	☐
Org_6. Reviews its mission, vision, and values with staff	☐	☐	☐	☐	☐
Org_7. Creates opportunities to share evaluation findings	☐	☐	☐	☐	☐
Org_8. Communicates about the importance of evaluation	☐	☐	☐	☐	☐
Org_9. Has resources for training staff for evaluation	☐	☐	☐	☐	☐
Org_10. Gives time to its staff for evaluation-related activities	☐	☐	☐	☐	☐
Org_11. Dedicates funds to conduct an evaluation	☐	☐	☐	☐	☐
Org_12. Has the technological capabilities (e.g., computers, software) to conduct an evaluation	☐	☐	☐	☐	☐
Org_13. Partners with external stakeholders/partners to help with the evaluation needs	☐	☐	☐	☐	☐
Org_14. Dedicates funds to ensure ongoing evaluation	☐	☐	☐	☐	☐
Org_15. Has resources in place to undertake evaluation on an ongoing basis	☐	☐	☐	☐	☐
Org_16. Integrates evaluation findings into decision-making	☐	☐	☐	☐	☐
Org_17. Involves external stakeholders in the evaluation process	☐	☐	☐	☐	☐
Org_18. Uses evaluation findings for external accountability	☐	☐	☐	☐	☐
Org_19. Uses evaluation findings for internal accountability	☐	☐	☐	☐	☐
Org_20. Has the commitment from external stakeholders to ensure evaluation sustainability	☐	☐	☐	☐	☐

B. Organizational capacity (Cont'd)

Based upon your own experience at work, please choose the best option for each statement.

The leadership of my organization…

	Strongly disagree	Disagree	Agree	Strongly agree	Don't Know
Org_21. Is open to new ideas	☐	☐	☐	☐	☐
Org_22. Is able to take risks in order to improve services	☐	☐	☐	☐	☐
Org_23. Builds ideas in collaboration with staff members	☐	☐	☐	☐	☐
Org_24. Resolves inter-personal conflicts in a positive manner	☐	☐	☐	☐	☐
Org_25. Encourages staff to be respectful of the needs of diverse populations (e.g., ethnicity, disability)	☐	☐	☐	☐	☐
Org_26. Celebrates staff members' achievements	☐	☐	☐	☐	☐
Org_27. Encourages learning new skills	☐	☐	☐	☐	☐
Org_28. Upholds organizational vision, mission, and goals	☐	☐	☐	☐	☐
Org_29. Promotes evaluative thinking	☐	☐	☐	☐	☐
Org_30. Is committed to evaluation	☐	☐	☐	☐	☐
Org_31. Has the skills necessary to undertake an evaluation	☐	☐	☐	☐	☐
Org_32. Encourages staff to reflect on evaluation findings	☐	☐	☐	☐	☐
Org_33. Shares evaluation findings with intended audience (e.g., partners, staff)	☐	☐	☐	☐	☐

C. Profile

The primary objective of this section is to gather some background information (e.g., age) about you and selected characteristics of your organization (e.g., type, size).

1. Are you:

Male ☐

Female ☐

Prefer not to specify ☐

2. Which of the following best describes your age range? Please √ one.

   29 years or less ☐

   30-39 years ☐

   40-49 years ☐

   50-59 years ☐

   60 years or more ☐

3. What is your highest level of education completed? Please √ one.

   Secondary (high) school diploma or equivalent ☐

   College, CEGEP, or other non-university certificate or diploma ☐

   University degree, certificate, or diploma ☐

   Graduate degree or higher (e.g., Masters) ☐

4. Which of the following **best** describes your position? Please √ one.

   Program director ☐

   Executive director/CEO ☐

   Manager/Supervisor ☐

   Administrator/support staff ☐

   Researcher/Academic ☐

   Educator ☐

   Evaluator ☐

   Volunteer ☐

   Other (Please specify) ☐

   _____

5. How long have you been with the present organization? Please √ one.

   Less than one year ☐

   1-4 years ☐

   5-9 years ☐

   10 years or more ☐

6. Is your organization a funding agency?

   Yes ☐

   No ☐

   Don't know /Not applicable ☐

7. Which of the following **best** describes your organization? Please √ one.

Government (local/provincial/federal) ☐

Academic (college/university) ☐

School board (public/private) ☐

Not-for-profit/community agency ☐

Foundation ☐

Other (Please specify) ☐

_____

8. Which of the following best describes the size of your organization? Please √ one.

9 or fewer individuals ☐

10-19 individuals ☐

20 or more individuals ☐

9. Where is your organization primarily located (i.e., the main office)? Please √ one.

Calgary ☐

Edmonton ☐

Grande Prairie ☐

Lethbridge ☐

Medicine Hat ☐

Fort McMurray ☐

Other (Please specify) ☐

_____

10. How many years of experience in evaluation do you have?

None ☐

Less than 1 year ☐

1 to 4 years ☐

5 years or more ☐

11. Have you ever taken professional development/training related to evaluation?

Yes ☐

No (Skip to Q14) ☐

12. Which of the following best describe the mode/source of the professional development/training? Please √ all that apply.

University/college degree/certificate ☐

Evaluation course ☐

Essential Skills Series course by the Canadian Evaluation Society ☐

On-site training ☐

Workshop ☐

Webinars ☐

Other (please specify): ☐

_____

13. Please comment on the usefulness of the professional development/training you have had:

_____
_____
_____

14. How was the professional development/training funded? Please √ one.

Current employer ☐

Past employer ☐

Personally funded ☐

Free of charge ☐

Other (please specify): ☐

_____

15. If you were given the opportunity to develop your evaluation knowledge and/or skills, which of the following would you prefer? Please √ all that apply.

University/college degree/certificate ☐

University/college level evaluation course ☐

Essential Skills Series course by the Canadian Evaluation Society ☐

On-site training ☐

Workshop ☐

Webinars ☐

Coaching or mentoring at work ☐

Community of practice (online) ☐

Other (please specify): ☐

_____

16. Please list the three most important reasons for your choice(s) above.

_____
_____
_____

Appendix E. Descriptive Statistics of Evaluation Capacity Items

This appendix includes two tables presenting descriptive statistical for individual and organizational EC.

1. Table 1: Descriptive statistical for Individual Evaluation Capacity

Items for Individual EC	N	Minimum	Maximum	Mean	Standard Deviation
Ind_1. I have an understanding of evaluation.	329	2	4	3.52	.52
Ind_2. I have a working knowledge of conducting an evaluation.	328	1	4	3.38	.64
Ind_3. I have a clear sense of the benefits of evaluation.	329	1	4	3.71	.51
Ind_4. I think evaluation is important for making improvements in my organization.	323	1	4	3.87	.36
Ind_5. I think evaluation can be useful in determining the effectiveness of my organization.	324	2	4	3.83	.38
Ind_6. I think evaluation findings can be beneficial to my sponsor(s)/funder(s).	39	3	4	3.78	.41
Ind_7. I think evaluation can improve transparency in an organization.	327	2	4	3.74	.45
Ind_8. I would be concerned if an organization doesn't evaluate its activities.	329	1	4	3.67	.56
Ind_9. I have an ethical responsibility to participate in evaluation as needed.	323	2	4	3.65	.51
Ind_10. I think it is important that staff members get involved in evaluation.	327	2	4	3.72	.48
Ind_11. I think evaluation creates unnecessary additional workload.	326	1	4	1.75	.67
Ind_12. I have the necessary resources (e.g., time, money) to engage in evaluation.	258	1	4	2.71	.75
Ind_13. I seek out professional development opportunities related to evaluation.	268	1	4	2.99	.71
Ind_14. I think time spent on evaluation is justified.	273	1	4	3.50	.54
Ind_15. I think internal allocation of funding for an evaluation is justified.	311	2	4	3.39	.55
Ind_16. I think credentialed training (e.g., course, certificate, diploma) in evaluation is important.	271	2	4	3.17	.65
Ind_17. I think non-credentialed training (e.g., webinar, coaching) in evaluation is important.	275	1	4	3.29	.52
Ind_18. I think external funding for an evaluation is justified.	313	2	4	3.36	.57

Ind_19. I know what skills to look for in an external evaluator.	259	1	4	2.66	.79
Ind_20. I am open to learning the necessary skills to conduct an evaluation.	320	1	4	3.42	.57
Ind_21. I am open to staff being provided with the opportunities to learn the skills necessary to conduct evaluation.	319	2	4	3.48	.52
Ind_22. I have the skills to oversee an external evaluator.	257	1	4	2.67	.85
Ind_23. I am open to adopting new ideas in my day-to-day activities based on evaluation findings.	323	2	4	3.60	.50
Ind_24. I know how to use evaluation findings in decision-making.	272	1	4	3.26	.61
Ind_25. I know how to make organizational level changes based on evaluation findings.	259	1	4	3.04	.71
Ind_26. I am open to adopting new practices or processes based on evaluation findings.	323	1	4	3.55	.52

2. Table 2: Descriptive statistical for Organizational Evaluation Capacity

Items for Organizational EC	*N*	Min	Max	Mean	Standard Deviation
My organization …					
Org_1. Encourages staff to express their opinions.	318	1	4	3.47	.66
Org_2. Involves staff when making long-term plans.	312	1	4	3.18	.81
Org_3. Gives staff the opportunity to reflect on organizational goals.	317	1	4	3.32	.73
Org_4. Provides professional development opportunities.	321	1	4	3.54	.68
Org_5. Provides opportunities for learning evaluation skills.	307	1	4	2.93	.78
Org_6. Reviews its mission, vision, and values with staff.	317	1	4	3.36	.70
Org_7. Creates opportunities to share evaluation findings.	309	1	4	3.17	.70
Org_8. Communicates about the importance of evaluation.	313	1	4	3.17	.76
Org_9. Has resources for training staff for evaluation.	297	1	4	2.67	.89
Org_10. Gives time to its staff for evaluation-related activities.	309	1	4	2.90	.76
Org_11. Dedicates funds to conduct an evaluation.	245	1	4	2.80	.76
Org_12. Has the technological capabilities (e.g., computers, software) to conduct an evaluation.	294	1	4	3.09	.74
Org_13. Partners with external stakeholders/partners to help with evaluation needs.	222	1	4	2.97	.78
Org_14. Dedicates funds to ensure ongoing evaluation.	214	1	4	2.83	.81
Org_15. Has resources in place to undertake evaluation on an ongoing basis.	230	1	4	2.87	.81
Org_16. Integrates evaluation findings into decision-making.	244	1	4	3.17	.65
Org_17. Involves external stakeholders in the evaluation process.	215	1	4	2.99	.73
Org_18. Uses evaluation findings for external accountability.	228	1	4	3.13	.70
Org_19. Uses evaluation findings for internal accountability.	283	1	4	3.14	.69
Org_20. Has the commitment from external stakeholders to ensure evaluation sustainability.	190	1	4	2.74	.84
The leadership of my organization …					
Org_21. Is open to new ideas.	318	1	4	3.40	.61
Org_22. Is able to take risks in order to improve services.	302	1	4	3.17	.70
Org_23. Builds ideas in collaboration with staff members.	314	1	4	3.30	.72
Org_24. Resolves inter-personal conflicts in a positive manner.	306	1	4	3.26	.71

Org_25. Encourages staff to be respectful of the needs of diverse populations (e.g., ethnicity, disability).	320	1	4	3.62	.56
Org_26. Celebrates staff members' achievements.	316	1	4	3.33	.69
Org_27. Encourages learning new skills.	317	1	4	3.52	.61
Org_28. Upholds organizational vision, mission, and goals.	317	1	4	3.51	.62
Org_29. Promotes evaluative thinking.	307	1	4	3.24	.72
Org_30. Is committed to evaluation.	258	1	4	3.29	.72
Org_31. Has the skills necessary to undertake an evaluation.	250	1	4	3.15	.74
Org_32. Encourages staff to reflect on evaluation findings.	256	1	4	3.18	.72
Org_33. Shares evaluation findings with intended audience (e.g., partners, staff).	259	1	4	3.24	.73