

Understanding and Measuring Evaluation Capacity: A Model and Instrument Validation Study

American Journal of Evaluation
34(2) 190-206
© The Author(s) 2013
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1098214012471421
aje.sagepub.com



Tina Taylor-Ritzler¹, Yolanda Suarez-Balcazar²,
Edurne Garcia-Iriarte³, David B. Henry², and
Fabricio E. Balcazar²

Abstract

This study describes the development and validation of the Evaluation Capacity Assessment Instrument (ECAI), a measure designed to assess evaluation capacity among staff of nonprofit organizations that is based on a synthesis model of evaluation capacity. One hundred and sixty-nine staff of nonprofit organizations completed the ECAI. The 68-item measure assessed participants' perceptions of individual and organizational predictors of two evaluation capacity outcomes: mainstreaming and use of evaluation findings. Confirmatory Factor Analysis and internal consistency results support the inclusion of the items and factors measured by the ECAI. Moreover, structural equation modeling results support the synthesis model and its depiction of relationships among evaluation capacity predictors and outcomes. We discuss the implications of using a validated model and instrument in evaluation capacity building research and practice.

Keywords

evaluation capacity building, evaluation capacity, measurement, instrument validation

Theoretical and practical interest in evaluation capacity building (ECB) is currently high, and building such capacity in nonprofit organizations is a focus of many evaluators (Carman, 2007; Preskill & Boyle, 2008). Yet, despite the prevalence of ECB activities and the publication of comprehensive ECB models (e.g., Nielsen, Lemire & Skov, 2011; Preskill & Boyle, 2008), scholars and practitioners alike assert that there is a need for: (1) more empirical research on the factors that comprise evaluation capacity and the relationships among these factors and (2) a validated instrument to both inform and measure the results of ECB efforts (Baizerman, Compton, & Stockdill, 2002; Carman

¹ Department of Psychology, Dominican University, River Forest, IL, USA

² University of Illinois at Chicago, Chicago, IL, USA

³ Trinity College Dublin, Dublin, Ireland

Corresponding Author:

Tina Taylor-Ritzler, Department of Psychology, Dominican University, 7900 W. Division Street, River Forest, IL 60305, USA.
Email: tritzler@dom.edu

2007; Carman & Fredericks, 2010; Compton, Baizerman & Stockdill, 2002; Compton, Glover-Kudon, Smith, & Avery, 2002; Cousins, Goh, & Elliott, 2007; Huffman, Thomas, & Lawrenz, 2008; King, 2002; Labin, Duffy, Meyers, Wandersman & Lesesne, 2012; Milstein, Chapel, Wetterhall, & Cotton, 2002; Preskill & Boyle, 2008; Stevenson, Florin, Mills, & Andrade, 2002; Stockdill, Baizerman, & Compton, 2002). Key questions for those involved in ECB efforts include *What factors are critical predictors of evaluation capacity outcomes? How are predictors interrelated? How can we assess evaluation capacity to inform ECB efforts within organizations? How can we measure the impacts of ECB processes within organizations?* To address these needs, we developed a synthesis model of evaluation capacity (see Suarez-Balcazar et al., 2010) and a resulting measure, the Evaluation Capacity Assessment Instrument (ECAI). The current study briefly describes the synthesis model and reports on the validation of the model and the ECAI.

Understanding Evaluation Capacity Building

ECB practices within nonprofit organizations are dynamic and complex, and are intended to make the use of evaluation processes routine (Cousins et al., 2007; Duffy, Labin, & Wandersman, 2007; Labin, Duffy, Meyers, & Wandersman, 2010; Stockdill et al., 2002). During the last decade, evaluators have increasingly used ECB practices to assist nonprofit organizations in developing evaluation capacity (Baizerman, Compton, & Stockdill, 2002; Carman & Fredericks, 2008; Compton, Glover-Kudon, Smith, & Avery, 2002; Garcia-Iriarte, Suarez-Balcazar & Taylor-Ritzler, 2011; Huffman et al., 2008; King, 2002; Labin et al., 2012; Milstein et al., 2002; Stevenson et al., 2002; Stockdill et al., 2002). ECB practices generally involve an evaluator providing training, technical assistance, consultation, and other activities to one or more staff within an organization or system (Duffy et al., 2007; Garcia-Iriarte et al., 2011). The goals of ECB practices are typically for staff within the target organization to document regularly the implementation of their programs, and to interpret their results, in order to understand and strengthen program implementation, improve program outcomes, and meet the accountability requirements of funders and accrediting bodies (Garcia-Iriarte et al., 2011; Gibbs, Napp, Jolly, Westover, & Uhl, 2002; King, 2002; Mackay, 2002; Milstein et al., 2002; Preskill & Boyle, 2008; Taut, 2007).

The relevant literature includes scholarship on conceptualizing (i.e., what is evaluation capacity?), building (i.e., how do you build evaluation capacity?), and measuring evaluation capacity (i.e., how do you measure evaluation capacity?). This literature includes a growing number of increasingly complex conceptual models of factors that impact ECB and evaluation capacity (e.g., Connolly & York, 2002; Cousins, Goh, Clark, & Lee, 2004; Nielsen et al., 2011; Preskill & Boyle, 2008), reports of ECB practices and case studies (e.g., Garcia-Iriarte et al., 2011; King & Volkov, 2005), and ECB and related assessment instruments (e.g., Botcheva et al., 2002; Cousins et al., 2007; Danseco, Halsall, & Kasprzak, 2009; Preskill & Torres, 2000; Taut, 2007; TCU-ORC, 2003; Volkov & King, 2007). Yet, despite the growing literature on building evaluation capacity, a dynamic and complex organizational process, the field lacks empirically validated models and corresponding assessment instruments that integrate and synthesize currently agreed upon components of evaluation capacity and allow for its measurement (Labin et al., 2012; Taylor-Ritzler, Suarez-Balcazar, & Garcia-Iriarte, 2009). Accordingly, the current study focuses on issues of conceptualization and measurement of evaluation capacity.

Evaluation Capacity: A Synthesis Model

To address the need for a common understanding of evaluation capacity, Suarez-Balcazar and her colleagues (2010) proposed a synthesis model that was developed through a systematic review of published conceptual models, ECB principles, and factors believed to sustain the practice of evaluation in nonprofit organizations, and over a decade of work with nonprofit organizations. They noted

that although evaluation scholars and practitioners differ in their emphasis on the relative importance and inclusion of factors that contribute to ECB practices, collectively they agree that the presence of both individual and organizational factors facilitates evaluation capacity, leading to evaluation mainstreaming and use among participating staff members (see Cousins, Goh, Clark, & Lee, 2004; Cousins, Goh, Elliott, & Aubry, 2008; Labin et al., 2012; Milstein & Cotton, 2000; Naccarella et al., 2007; Nielsen et al., 2011; Preskill & Boyle, 2008; Stockdill et al., 2002; Volkov & King, 2007). Suarez-Balcazar et al.'s (2010) synthesis model distills the complexity and diversity of factors that have been identified by scholars and practitioners into a single, unified, and parsimonious model.

This synthesis model includes individual and organizational factors that are believed to predict evaluation capacity outcomes as well as the outcomes themselves. These are (a) Individual factors that contribute to evaluation capacity, including Awareness of the benefits of evaluation, Motivation to conduct evaluation, and Competence (knowledge and skills) to engage in evaluation practices; (b) Organizational factors that contribute to evaluation capacity, including Leadership for evaluation, a Learning Climate that fosters evaluative thinking, and Resources that support evaluation; and (c) critical Evaluation Capacity Outcomes, including Mainstreaming evaluation practices into work processes and Use of evaluation findings (see Suarez-Balcazar et al., 2010, for a review of the ECB literature and description of the development and components of the model).

Moreover, the Suarez-Balcazar et al. (2010) model moves beyond identifying predictors of evaluation capacity to specify how individual and organizational factors relate to evaluation capacity outcomes. Based on the authors' and others' work (e.g., Carman & Fredericks, 2010; Cousins et al., 2004; Garcia-Iriarte et al., 2011; Nielsen et al., 2011), the model posits that organizational factors affect the relationship between individual factors and evaluation capacity outcomes. Specifically, ECB practices are intended to develop individual knowledge, skills, and attitudes related to evaluation among the organization's staff. In turn, organizational learning capacity, which includes leadership, culture, systems and structures, and communication, is assumed to facilitate and/or hinder the transfer of individual learning into organizational processes and practices. It is expected that the presence of favorable organizational learning capacity conditions such as effective leadership, a learning climate, and adequate resources will be related to sustainable evaluation practice, defined as continuous learning about evaluation and use of evaluation findings, frameworks, and processes. This understanding of evaluation capacity is consistent with that of other scholars (e.g., Preskill & Boyle, 2008) and suggests that the relationship between individual factors and sustainable evaluation practice (evaluation capacity outcome) is mediated by an organization's commitment to developing evaluation capacity and its ability to change (organizational factors).

The ECAI: A Synthesis Measure

The decision to empirically test the synthesis model led to the need to develop an assessment instrument that measures the factors included in the model. Measuring evaluation capacity has been a challenge for evaluation scholars and practitioners. The majority of current instruments were developed from case studies and systematic analyses of the literature (e.g., Danesco et al., 2009; Preskill & Torres, 2000; Volkov & King, 2007); none were designed to validate empirically a conceptual model of evaluation capacity, and only a few provide psychometric data. Specifically, three measures report descriptive statistics and measurement of the latent construct of evaluation capacity (see Botcheva et al., 2002; Cousins et al., 2008; TCU-ORC, 2005). Careful review of the descriptions of these measures reveals that, collectively, they assess a wide variety of individual and organizational factors. To integrate current measurement, we developed the ECAI based on the Suarez-Balcazar et al. synthesis model and a review of publicly available ECB measures (see Table 1 for a summary). The current study empirically validates

Table 1. ECB Assessment Instruments.

Name of Instrument	Author and Year	Components Measured by the Instrument	Number of Items	Validation/Replication of Instrument
Assessing Learning Culture scale	Boccheva, White, and Huffman (2002)	Outcome measurement practices Learning culture	10	Developed based on the aspects of learning culture outlined by Preskill and Torres (2000). Empirical evidence of the relationship among outcomes measurement practices, learning culture, and external funding are shown. High reliability and internal consistency are described in a study with 25 agencies
Evaluation and organizational capacity	Cousins, Goh, Elliot, and Aubry (2008)	Organizational learning capacity Organizational support systems Capacity to do evaluation Specific types of evaluation activities Stakeholder participation in evaluation Use of evaluation findings Use of evaluation process Conditions mediating evaluation use Experience with evaluation Leadership and collaboration Systems and structures Evaluation practice Culture (organizational) Leadership Systems and structures Communication of information Teams (working as a team) Evaluation	120	Developed as a hybrid instrument to capture the constructs of interest. Based on existing and new scales. Data from 340 respondents were used to provide empirical evidence for the variables that explain evaluation in the voluntary sector
Readiness assessment tool for evaluation capacity building	Daneseo, Halsall, and Kasprzak (2009)		26	No validation and/or psychometric data are available
Readiness for Organizational Learning and Evaluation (ROLE)	Preskill and Torres (2000)	Culture (organizational) Leadership Systems and structures Communication of information Teams (working as a team) Evaluation	78	Developed based on a literature review to identify themes and patterns of evaluation capacity building. No validation and/or psychometric data are available
Evaluation process use measure	Taut (2007)	Section 1: Views of evaluation, decision making, expectations, sharing knowledge, and learning culture Section 2: Opinions and experiences with evaluation, available resources, internal and external monitoring, and reporting Section 3: Previous experience with evaluation	69	No validation and/or psychometric data are available
Organizational readiness for change (TCU-ORC)	TCU Institute of Behavioral Research (2005)	Motivation for change (program needs, training needs, pressure for change) Resources Staff attributes Organizational climate Organizational Context ECB structures Resources		Developed as part of an organizational research program at Texas Christian University. Used a national sample of over 500 staff members from more than 100 programs to assess reliability and validity of the ORC
A checklist for building organizational evaluation capacity	Volkov and King (2007)	Organizational Context ECB structures Resources	35	Developed from case study data and a literature review. No validation and/or psychometric data are available

the ECAI, assessing the specific items and individual and organizational factors that comprise evaluation capacity.

In sum, the field lacks a validated, generalizable model and instrument to assess the predictors of evaluation capacity and the degree to which it is developed and sustained. We found no studies that empirically assess whether and how individual and organizational factors are related. Specifically, no studies have tested the presence or degree of interaction between individual and organizational factors on evaluation capacity outcomes. As Preskill and Boyle (2008) have noted, more research is needed to guide the conceptualization and measurement of factors that are related to evaluation capacity.

Purpose of the Current Study

Using quantitative methods, we sought to test: (1) the validity of the ECAI by assessing the multiple proposed factors that are related to evaluation capacity and the items that are included on each factor in the instrument; and (2) the Suarez-Balcazar et al. (2010) synthesis model by examining the degree to which the organizational factors of leadership, resources, and learning climate affect the relationship of the individual factors of awareness, motivation, and competence to the evaluation capacity outcomes of mainstreaming and use. Thus, we wish to add an empirical validation study to the evaluation literature and make the ECAI available for use by evaluation scholars and practitioners who seek to diagnose and build evaluation capacity within organizations.

Method

Instrument: The Evaluation Capacity Assessment Instrument

The development of the ECAI followed five established procedures in measurement development (see Clark & Watson, 1995). First, we conducted a search of the literature to identify existing measures of evaluation capacity. Second, we reviewed the literature and proposed a set of indicators to measure the factors in the Suarez-Balcazar et al. (2010) model: (a) Individual Factors, including Awareness of the benefits of evaluation, Motivation to conduct evaluation, and Competence (knowledge and skills) to engage in evaluation practices; (b) Organizational Factors, including Leadership for evaluation, a Learning Climate that fosters evaluative thinking, and Resources that support evaluation; and (c) critical Evaluation Capacity Outcomes, including Mainstreaming evaluation practices into work processes and Use of evaluation findings.

Third, we identified and developed items to measure indicators for each of the constructs. We developed 56 items and adapted 12 from other instruments. Specifically, 6 items were from Preskill and Torres' (2000) Readiness For Organizational Learning and Evaluation (ROLE) questionnaire (two were used verbatim and four were reworded); 4 items were adapted from the Texas Christian University's Survey of Organizational Functioning (2005; 2 items were used verbatim and two were reworded); and 2 items were reworded from Taut (2007). The ECAI instrument included a total of 68 items. Items on the ECAI used 4-point response scales (where 4 = *strongly agree*, 3 = *somewhat agree*, 2 = *somewhat disagree*, and 1 = *strongly disagree* or 4 = *to a very great extent*, 3 = *to a considerable extent*, 2 = *to some extent*, and 1 = *not at all*). The ECAI also included demographic questions that asked participants to indicate their gender, race/ethnicity, position within the organization (e.g., Executive Director, Administrator/Manager, Service Worker, etc.), educational attainment, prior experiences with evaluation, number of years employed in current position, and their organization's funding sources.

Fourth, a panel of experts composed of three experienced national evaluation scholars reviewed all of the items and provided feedback. The panel included two academics who regularly teach evaluation courses, routinely conduct evaluations, and have authored evaluation reports and/or a widely used evaluation text, and one member who has been a senior evaluator at a large evaluation

organization and is a national evaluation consultant. We asked the panel of experts to focus on the content of the instrument, the wording of each item, the uniqueness of each item compared to other items in the instrument, and the subfactor/factor that the item referenced. The reviewers' feedback was used to clarify confusing items, delete repeated items, and develop new items.

Fifth, we pilot tested the ECAI with four staff members from different nonprofit organizations. Pilot participants completed the instrument and provided feedback on items that were confusing, difficult to understand, or difficult to rate, and gave their overall impression of the content, format, and length of the survey. Their feedback was incorporated into the revision of the instrument.¹

Sample Recruitment

We obtained names and addresses of nonprofit organizations in the Chicago metropolitan area from the Community Resource Network's list of nonprofits in Illinois, which includes several thousand primarily social service agencies. Organizations varied from small, grassroots, and volunteer driven to large and multisite. We specified a geographic band around the city of Chicago and suburban communities that resulted in a list of 1,021 nonprofit organizations and then randomly sampled 1,000 from this list due to a budget of only \$1,000 for incentives. We invited 1,000 organizations to participate in order to obtain a large enough sample to statistically validate our model and instrument. We mailed an envelope to the attention of the executive director of each organization that included a cover letter, a copy of the survey, a stamped self-addressed return envelope, and a one dollar bill as an incentive to participate. The cover letter included a link to an online version of the survey. After 4 weeks, we sent a reminder postcard to each nonrespondent and also included the link to the online version of the survey. One hundred and fifty-one surveys were returned with an incorrect address, resulting in a total target sample of 849. Of the 849 surveys that are presumed to have reached their intended recipient, 169 were completed and returned (155 hard copy and 14 online), resulting in a response rate of 20%.

Study Participants

Respondents reported that they were executive directors or administrators (53%), managers (35%), or service workers/clinical staff (11%). They had worked in their current position for an average of 8.65 years ($SD = 7.63$; range: less than 1 year to 38 years). In terms of educational attainment, 69.4% had a master's degree or higher, 21.3% had a bachelor's degree, and 7.1% had less than a bachelor's degree. The majority of respondents were female (64%). Respondents identified themselves as Caucasian (52%), African American (30%), Latino/Latina (13%), or Asian American (3%). Funding information was reported by 69% of the respondents. Of this group, 49% reported that they had funding from foundations, while 32% had funding from other sources, including government, universities, churches, and associations; 19% reported receiving funding from both foundations and other sources.

Seventy four percent of the respondents had evaluation experience. Of these, 24% reported one experience and 50% reported two or more. These experiences included participation in an actual evaluation (50% of respondents), completion of an evaluation course (43%), attendance at an evaluation workshop (48%), or attendance at one or more evaluation presentations (46%). Thirty-eight percent of respondents indicated that their organization did have an evaluation department or one or more evaluators on staff. Organizations with evaluators employed an average of three individuals ($SD = 5.53$, range: a quarter-time position to 30 people).

Data Analysis: Instrument and Model Validation

Due to the modest sample size, we analyzed the data using a two-step process that included both confirmatory factor analysis (CFA) and SEM (Jöreskog, 1993). Our sample was adequate for these

analyses, as K. G. Jöreskog (personal communication, June 1994) has indicated no absolute minimum sample size-to-parameter ratio for SEM analyses. In Step 1, we tested the validity of the ECAI using three separate CFAs to map the individual items onto the theorized first-order latent constructs that were indicators of the second-order latent constructs. Specifically, the three CFA analyses modeled the relationships of: (1) manifest indicators on Awareness (11 items), Motivation (4 items) and Competence (14 items), which served as first-order latent constructs, on a second-order latent construct called Individual Factors; (2) manifest indicators on Leadership (5 items), Learning Climate (9 items), and Resources (9 items), which served as first-order latent constructs, on a second-order latent construct called Organizational Factors; and (3) manifest indicators on Mainstreaming (5 items) and Use (11 items), which served as first-order latent constructs, on a second-order latent construct called ECB Outcomes. Residual correlations among the items were allowed only if similarities in wording justified them. Correlations between the first order factors were allowed.

In Step 2, we validated the synthesis model by computing mean composite scores for each of the eight first-order latent constructs (awareness, motivation, competence, leadership, learning climate, resources, mainstreaming, and use) and using SEM to model the relationships between the latent individual and organizational predictor factors and the outcomes. In so doing, we tested the degree to which organizational factors mediated the relationship between individual factors and outcomes (Jöreskog, 1993). This dual-step analytic process used procedures that have been reported in studies using SEM (see Gorman-Smith, Tolan, Zelli, & Huesmann, 1996; Tolan, Gorman-Smith, & Henry, 2003).

In this validation process, we used four fit indices to evaluate the results of the three CFAs and one SEM. These fit indices were relative chi-square (χ^2/df ; Carmines & McIver, 1981); root mean square error of approximation (RMSEA, Browne & Cudeck, 1992), comparative fit index (CFI, Bentler, 1990), and Tucker–Lewis index (TLI, Bentler & Bonett, 1980). Based on established guidelines, we adopted specific cutoff values for each fit index (Byrne, 1991; Hu & Bentler, 1999). Specifically, the relative chi-square and RMSEA values should be lower than 2 and .06, respectively; CFI and TLI indices should be at least .90; and factors should be both highly interpretable and parsimonious (see Bentler & Mooijaart, 1989). After final factor solutions were obtained, the internal consistency (i.e., Cronbach's α) of each of the factors was computed, with values of .80 or higher determined to be good (Nunnally & Bernstein, 1994).

We conducted all analyses using Mplus Version 5.21 (Muthén & Muthén, 2009). We estimated the parameters using the maximum likelihood method (MLM) and obtained reliability estimates. Mplus uses full information maximum likelihood (FIML) estimation, which permits use of the entire sample ($N = 169$) for all analyses. On each variable, more than 95% of the data was nonmissing. Without FIML, 22 to 33 cases would have been removed from analysis.

Results

In Table 2, we report fit indices for the three CFAs and the SEM. In Table 3, we report the descriptive statistics and internal consistencies for each study construct. In Figure 1, we present diagrams of the CFAs and SEM.

Step 1: CFA

As noted, the first purpose of this study was to validate the ECAI by assessing the relationship between the items and first- and second-order latent constructs. Overall, results showed that the eight first-order and two of the three second-order CFA models were an acceptable fit for the data. Specifically, results showed that for the second-order latent construct Individual Factors, all fit

Table 2. Indices of Fit.

Model	χ^2	df	χ^2/df	RMSEA	CFI	TLI
Step 1: CFA results						
CFA #1: Individual factors	525.20*	367	1.43	.059	.917	.908
CFA #2: Organizational factors	284.80*	223	1.28	.043	.950	.943
CFA #3a: Evaluation capacity outcomes ^a	311.20*	102	3.05	.118	.842	.814
CFA #3b: Mainstreaming & use ^b	151.32*	100	1.51	.059	.961	.953
Step 2: SEM						
Structural model	16.83	12	1.40	.049	.990	.976

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index.

* $p < .05$.

^aEvaluation capacity outcomes was not a good fit for the data.

^bMainstreaming and use were maintained as separate first-order factors. They were correlated, $\beta = .712$, $p < .05$.

Table 3. Descriptive and Internal Consistency Statistics of Study Constructs.

Variable	# of Items	M	Range	SD	α
Individual factors	29	3.15	1–4	.39	.93
Awareness	11	3.28	1–4	.45	.87
Motivation	4	3.15	1–4	.57	.90
Competence	14	3.01	1–4	.48	.94
Organizational factors	23	2.91	1–4	.36	.90
Leadership	5	3.07	1–4	.39	.82
Learning climate	9	3.10	1–4	.43	.86
Resources	9	2.57	1–4	.55	.88
Outcomes	—	—	—	—	—
Mainstreaming	5	2.90	1–4	.53	.87
Use	11	2.78	1–4	.68	.92

indices met the fit criteria established a priori. Results also indicated that for the second-order latent construct Organizational Factors, all fit indices met our criteria for model fit.

Results of the two-component second-order latent construct Evaluation Capacity Outcomes indicated that the model was not a good fit for the data (see Step 1: CFA #3a). An alternate model, in which Mainstreaming and Use were maintained as separate first-order latent constructs and not combined into a second-order construct, was a better fit (see Step 1: CFA #3b). Fit indices for this alternate model met the established criteria of merit. Taken together, the items included in the ECAI tap into the latent constructs as predicted. In addition, internal consistency analyses revealed that the subfactors and factors demonstrated strong reliability (see Table 3).

Step 2: SEM

A structural model was tested to better understand the relationship between individual and organizational factors and the evaluation capacity outcomes of mainstreaming and use. As noted, we computed mean composite scores for each of the first-order latent constructs and conducted several analyses using these means. As a result, the first-order latent constructs of Awareness, Motivation, Competence, Leadership, Learning Climate, Resources, Mainstreaming and Use that were tested in Step 1 became manifest constructs in Step 2. Before conducting the SEM, correlations among the mean composite scores on the factors were computed (see Table 4). Results showed that both the Individual and Organizational Factors were significantly, positively and generally moderately

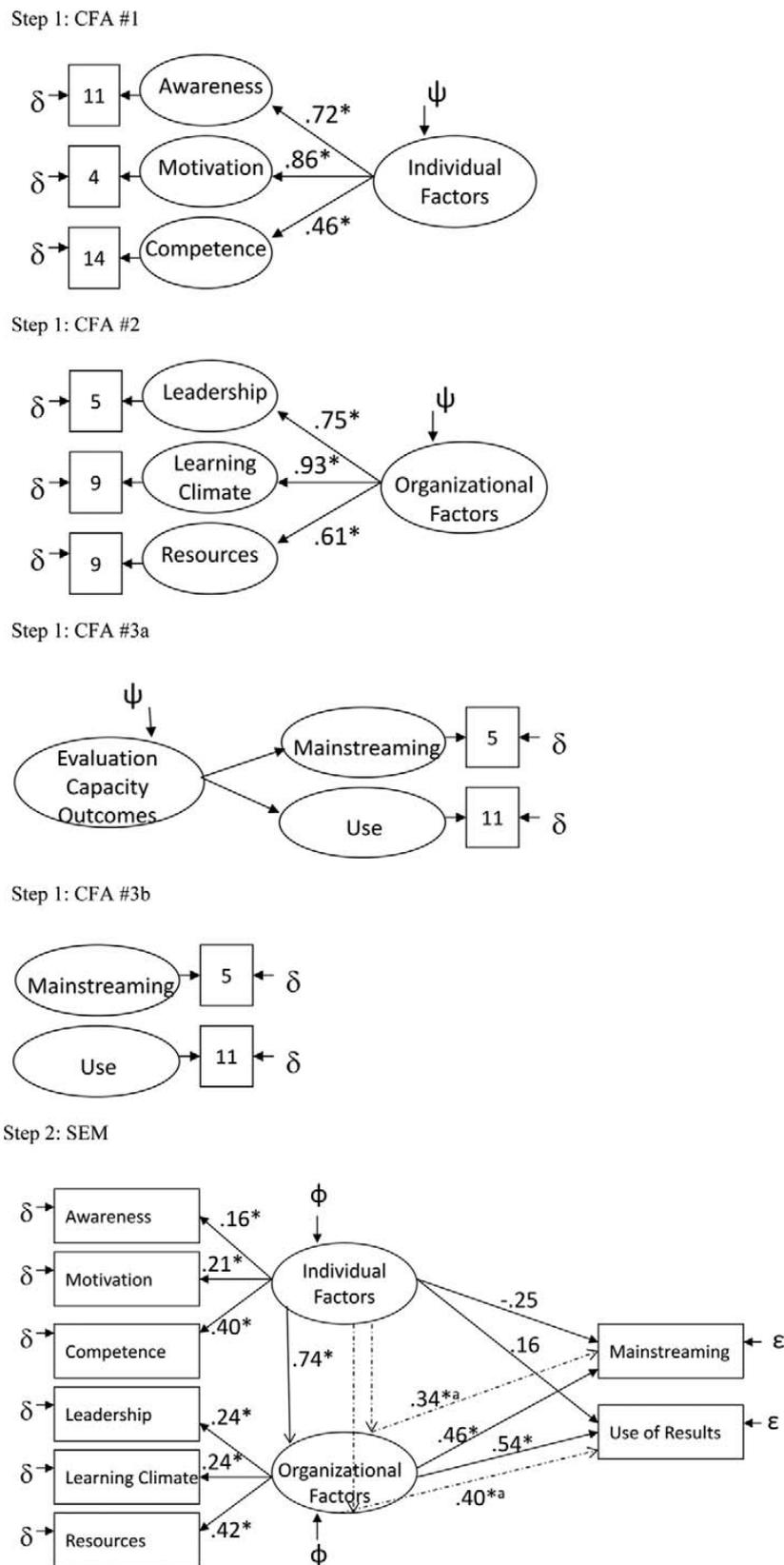


Figure 1. Confirmatory factor analysis and structural equation model results.
 Note. * $p < .01$. Solid lines represent direct effects. Dashed lines and ^a represent indirect effects. Parameter estimates are reported. δ refers to errors of “X”-side manifest variables. ψ refers to errors of “Y”-side latent variables. ϕ refers to errors of “X”-side latent variables. ϵ refers to errors of “Y”-side manifest variables. The model in Figure 3a was not a good fit for the data. The model in Figure 3b was a good fit, with mainstreaming and use maintained as separate factors that did not load on the higher order latent variable of *Evaluation Capacity Outcomes*.

Table 4. Correlations Matrix of Study Constructs.

	1	2	3	4	5	6	7	8
1. Awareness	—	—	—	—	—	—	—	—
2. Motivation	.55*	—	—	—	—	—	—	—
3. Competence	.32*	.32*	—	—	—	—	—	—
4. Leadership	.06	.07	.36*	—	—	—	—	—
5. Learning climate	.29*	.22*	.50*	.55*	—	—	—	—
6. Resources	.06	.17*	.45*	.45*	.40*	—	—	—
7. Mainstreaming	.25*	.25*	.52*	.56*	.57*	.67*	—	—
8. Use	.32*	.25*	.50*	.43*	.46*	.57*	.71*	—

* $p < .05$.

intercorrelated and correlated with Mainstreaming and Use. We also screened the data for univariate outliers (see Table 3) and used MLM estimation that is robust to deviations from multivariate normality.

The composite variables were used in the covariance structure model that is depicted in Step 2 of Figure 1. The model shows the relationships among the latent constructs, including indirect effects testing in Mplus. Indices of fit for the SEM are presented in Table 2 and indicate that the proposed model was a good fit to the data. The model regressed Mainstreaming and Use on Organizational Factors and Organizational Factors on Individual Factors. Thus, in Step 2 Individual Factors was an exogenous latent variable with three manifest indicators, Organizational Factors was an endogenous latent variable with three manifest indicators, and Mainstreaming and Use were endogenous manifest variables. We also fit an alternative model that reversed the indirect path from Individual Factors to the outcomes via Organizational Factors. Both models had 12 degrees of freedom, so a likelihood ratio test comparison was not possible. However, the alternative model had a significant chi-square goodness-of-fit test, $\chi^2(12, N = 169) = 21.18, p < .05$, whereas the final model had a nonsignificant test, $\chi^2(12, N = 169) = 16.83, ns$.

The correlations in Table 4 show that the components of Individual and Organizational Factors were all significant positive predictors of Mainstreaming and Use, suggesting that higher levels of both Individual Factors (awareness, motivation, and competence) and Organizational Factors (leadership, learning climate, and resources) were related to higher levels of Mainstreaming and Use. However, when both individual and organizational factors are modeled in relationship to Mainstreaming and Use, the best fitting model had relationships between Individual Factors and Mainstreaming and Use fully mediated by Organizational Factors.

Discussion

The current study was designed to fill a gap in the evaluation literature by reporting on the validation of an evaluation capacity synthesis model and the new ECAI, making this tool available for use by both evaluation scholars and practitioners. SEM results provide empirical evidence to support Suarez-Balcazar et al.'s (2010) synthesis model. CFA results support the use of the multicomponent ECAI as a generalizable tool for measuring evaluation capacity in nonprofit organizations. These findings add to the research base related to the conceptualization and measurement of evaluation capacity, and as such, may contribute to efforts to build evaluation capacity within organizations.

The results are in line with past research and strengthen a conceptual foundation for understanding evaluation capacity. The evidence is further strengthened by the use of statistical techniques that

advance our understanding of evaluation capacity predictors and their relationship to outcomes. The good fit of the eight first-order and two second-order latent factors suggests that the constructs in the model proposed by Suarez-Balcazar et al. (2010) are a viable way of understanding evaluation capacity.

Findings support the hypothesized relationship between individual and organizational factors and evaluation capacity outcomes. Specifically, organizational factors play a crucial role in the relationship between individual factors and the mainstreaming of evaluation practices into the day-to-day work of organizations, and also using evaluation results to understand and improve programs. Others have indicated that organizational factors are critical in ECB processes (Cousins et al., 2004; Preskill & Boyle, 2008). The current study provides statistical evidence of the importance of organizational factors in understanding evaluation capacity. Other researchers (Cousins et al., 2004; Preskill & Boyle, 2008) have indicated that the contextual embeddedness of organizational factors impact the development of evaluation capacity among individual staff. Our results support what many who practice ECB have observed: Even when individual staff members have the knowledge and motivation to engage in evaluation activities such as mainstreaming and use, these activities are less likely to occur if their organization does not provide the leadership, support, resources, and necessary learning climate.

Limitations of the Current Study and Future Research

Certain limitations should be considered when interpreting our findings. The first set of limitations is related to the study sample. The response rate was low; the sample was drawn exclusively from one area of the United States and the sample size was relatively small given the number of estimated parameters. Also, roughly half of the respondents were executive directors of the organizations sampled. No information was collected about the size and sophistication of the organization. Thus, the results should be interpreted with caution. We recognize that the perspectives of organizational leaders on organizational evaluation capacity might not represent the perspectives of staff who implement programs and deal with the challenges of incorporating evaluation strategies within their daily service activities. Views of evaluation likely differ among staff within an organization depending on their position and roles. Our survey included forced response options for the question about the respondents' role within the organization and it is not clear from our data whether, and to what extent, respondents were the individuals within their organizations who were ultimately responsible for evaluation activities. The current study did not assess the complexity of the nonprofit organizations in the sample, including their size, capacity, and context, nor did it assess the role of evaluators within the organization. These are all important contextual factors that are likely to impact evaluation capacity and should be considered in future research to assess whether the model and measure are a good fit for data with new samples.

The second set of limitations is related to methodology. This study should be considered exploratory, given that the same sample was used for both the CFA and SEM analyses, and that the most appropriate design would have been a longitudinal multilevel modeling approach. Although there is precedent in the literature for the dual-step analytic process used in the current study (see Gorman-Smith et al., 1996; Tolan et al., 2003), there is a risk of distorted effects due to sampling error associated with using the same sample for both sets of analyses. Future research should replicate the current study with an independent sample. Moreover, the ideal research design would involve sampling multiple participants from different organizations over time in order to estimate both within- and between-organization variance, and to make claims about the temporal precedence of the predictor and outcome variables. Longitudinal assessment would allow for measuring the degree to which ECB efforts result in improvements in evaluation capacity and the degree to which these are sustained over time.

The strong path coefficient between the latent variables for Individual and Organizational factors as well as the moderate to high correlations between their indicator variables suggest that multicollinearity may have reduced the extent to which the effects of Individual and Organizational factors on Mainstreaming and Use of Results could be differentiated. This possibility is rendered less likely by the fact that the model as depicted in Figure 1 fits the data better than the alternative model that reversed the direction of the mediated effect.

The third set of limitations is related to the simplicity of the model that was tested. The Saurez-Balcazar et al. model is, by design, parsimonious. As a result, it does not include all of the important indicators of evaluation capacity that have been identified in the literature. For example, process use is not included in the current model. Important organizational support systems (e.g., incentives and rewards), internal pressures from program participants and staff, and external pressures from funders and accreditation requirements are not included. Cultural factors related to the organization, the program, and the participants the organization serves should also be considered in future research. Taken together, these investigations could increase the complexity of the model and the ECAI by incorporating additional constructs and contextual elements that may impact evaluation capacity.

Implications for Evaluation Practice

The synthesis model of evaluation capacity and ECAI in their current form could be used by program evaluators to guide ECB efforts. Specifically, in conjunction with the model, the ECAI may inform practitioners about where to place their emphasis when conducting ECB activities. Practitioners could employ the ECAI to conduct pretest assessments of organizational evaluation capacity and use the results to focus their ECB efforts. As such, we recommend the ECAI as a diagnostic tool for evaluation planning, training, consultation, and technical assistance purposes. In contexts where the ECAI may not be appropriate (e.g., where literacy levels are low or where surveys are not part of the organization's norms), the validated model, without use of the ECAI, may be informative in guiding other types of assessment efforts to gauge evaluation capacity (i.e., observational or qualitative approaches).

Our results suggest that ECB efforts should focus on organizational leadership, a learning culture, and developing appropriate and adequate resources devoted to evaluation. This represents a shift in emphasis, as ECB efforts have traditionally targeted individual factors such as motivation and skills. In addition, the ECAI or other measures could be used after ECB efforts are completed to assess changes in organizational capacity that reflect outcomes of ECB efforts on the individual, organizational, and outcome factors that are included in the model. The synthesis model and ECAI could also be used over time to assess the sustainability of evaluation capacity within organizations and inform ongoing ECB efforts.

Conclusion

The current study informs our understanding of, and ability to measure, evaluation capacity. It identifies organizational and individual factors associated with evaluation capacity outcomes, and advances our analysis of the relationship between these factors and their role in evaluation mainstreaming and use. The findings also provide a foundation for building empirical evidence that supports the conceptualization and measurement of evaluation capacity and guides efforts to enhance ECB implementation and evaluation. Future research is needed to replicate this study, add to our understanding of complex contextual factors, and assess evaluation capacity across a variety of organizations and across time within the same organization.

Appendix

Items on the Evaluation Capacity Assessment Instrument (ECAI)

Section I: About You (Individual Factors) Awareness: Thoughts About Evaluation¹ I think that an evaluation . . .

1. Will help me understand my program.
2. Will inform the decisions I make about my program.
3. Will justify funding for my program.
4. Will help to convince managers that changes are needed in my program.
5. Will inform changes in our documentation systems.
6. Is absolutely necessary to improve my program.
7. Should involve program participants in the evaluation process.
8. Will influence policy relevant to my program.
9. Will help improve services to people from diverse ethnic backgrounds who also have disabilities
10. Is unnecessary because we already know what is best for our participants.
11. Is too complex for staff to do.

Motivation: Motivation to Engage in Evaluation¹

I am motivated to . . .

1. Learn about evaluation.
2. Start evaluating my program.
3. Support other staff to evaluate their program.
4. Encourage others to buy into evaluating our program.

Competence: Evaluation Knowledge and Skills¹

I know how to . . .

1. Develop an evaluation plan.
2. Clearly state measurable goals and objectives for my program.
3. Identify strategies to collect information from participants.
4. Define outcome indicators of my program.
5. Decide what questions to answer in an evaluation.
6. Decide from whom to collect the information.
7. Collect evaluation information.
8. Analyze evaluation information.
9. Develop recommendations based on evaluation results.
10. Examine the impact of my program on people from diverse ethnic/racial backgrounds and/or people with disabilities.
11. Write an evaluation report.
12. Conduct an evaluation of my program on my own.
13. Conduct an evaluation of my program with support from others.
14. Present evaluation findings orally.

Section II: About your Organization (Organizational Factors)

Leadership¹

1. Program managers provide effective leadership.
2. Staff understands how everyone's duties fit together as part of the overall mission of the program.

3. Program managers communicate program goals and objectives clearly.
4. Program managers have a clear plan for accomplishing program goals.
5. Program managers have realistic expectations of what staff can accomplish given the resources they have available.

Learning Climate¹

The program where I work fosters an environment in which . . .

1. Evaluation information is shared in open forums.
2. Staff is supported to introduce new approaches in the course of their work.
3. It is easy for staff to meet regularly to discuss issues.
4. Staff is provided opportunities to assess how well they are doing, what they can do better, and what is working.
5. Staff can encourage managers and peers to make use of evaluation findings.
6. Staff respects each other's perspectives and opinions.
7. Staff errors lead to teachable moments rather than criticisms.
8. Staff participates in making long-term plans for their program.
9. Staff concerns are ignored in most decisions regarding strategic planning and evaluation.

Resources for Evaluation¹

In my program . . .

1. Resources are allocated to provide accommodations for people from diverse ethnic backgrounds and for people with disabilities to collect evaluation information (e.g., interpreters, translated documents).
2. Staff has time to conduct evaluation activities (e.g., identifying or developing a survey, collecting information from participants).
3. Staff has access to technology to compile information into computerized records.
4. Staff has access to adequate technology to produce summary reports of information collected from participants (e.g., computerized database).
5. Resources are allocated for staff training (e.g., money, time, bringing in consultants).
6. Technical assistance is available to staff to address questions related to evaluation.
7. Funders provide resources (e.g., training, money, etc.) to conduct evaluation.
8. Funders provide leadership for conducting evaluation.
9. Agency leadership engages in ongoing dialogue with funders regarding evaluation.

Section III: About your Work (Evaluation Capacity Outcomes)

Mainstreaming: Evaluation as part of your Job¹

1. My program gathers information from diverse stakeholders to gauge how well the program is doing.
2. My program has adequate records of past evaluation efforts and what happened as a result.
3. I have access to the information I need to make decisions regarding my work.
4. I am able to integrate evaluation activities into my daily work practices.
5. The evaluation activities I engage in are consistent with funders' expectations.

Use of Evaluation Findings²

Please indicate the extent to which your program currently uses evaluation results for the following purposes

1. To report to a funder.
2. To improve services or programs.

3. To get additional funding.
4. To design ongoing monitoring processes.
5. To assess implementation of a program.
6. To assess quality of a program.
7. To improve outreach.
8. To make informed decisions.
9. To train staff.
10. To develop best practices.
11. To eliminate unneeded services or programs.

Note. 1. Response format was a 1-4 scale, where 1 = strongly disagree, 2 = somewhat disagree, 3 = somewhat agree, and 4 = strongly agree.

2. Response format was a 1-4 scale, where 1 = *not at all*, 2 = *to some extent*, 3 = *to a considerable extent*, and 4 = *to a very great extent*.

Authors' Note

The opinions expressed herein are those of the authors and not necessarily those of National Institute on Disability and Rehabilitation Research (NIDRR). The authors wish to acknowledge the valuable contributions of Rooshey Hasnain, Asma Ali, Juleen Rodakowski, and Lucia Orellana-Damacela.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This study was funded by the NIDRR, U.S. Department of Education, through a grant to the Center for Capacity Building on Minorities with Disabilities Research at the University of Illinois at Chicago (Grant # H133A040007).

Note

1. The items in the instrument are included in the Appendix.

References

- Baizerman, M., Compton, D. W., & Stockdill, S. H. (2002). New directions for ECB. *New Directions for Evaluation*, 93, 109-119.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin*, 106, 315-317.
- Botcheva, L., White, C. R., & Huffman, L. C. (2002). Learning culture and outcome measurement practices in community agencies. *American Journal of Evaluation*, 23, 421-434.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230-258.
- Byrne, B. M. (1991). The Maslach Burnout Inventory: Validating factorial structure and invariance across intermediate, secondary, and university educators. *Multivariate Behavioral Research*, 26, 583-605.

- Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation, 28*, 60–75.
- Carman, J. C., & Fredericks, K. A. (2008). Nonprofits and evaluation: Empirical evidence from the field. In J. G. Carman & K. A. Fredericks (Eds.), *Nonprofits and evaluation. New directions for evaluation (Vol. 119, pp. 51–71)*. San Francisco, CA: Jossey-Bass.
- 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*, 84–104.
- Carmines, E., & McIver, J. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. Bohrnstedt & E. Borgatta (Eds.), *Social measurement: Current issues* (pp. 65–115). Beverly Hills, CA: Sage.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issue in objective scale development. *Psychological Assessment, 7*, 309–319.
- Compton, D., Baizerman, M., & Stockdill, S. (2002). Special Issue: The art, craft, and science of evaluation capacity building. *New Directions for Evaluation, 93*, 1–120.
- Compton, D., Glover-Kudon, R., Smith, I. E., & Avery, M. E. (2002). Ongoing capacity building in the American Cancer Society (ACS) 1995–2001. The art, craft, and science of evaluation capacity building. *New Directions for Evaluation, 93*, 47–61.
- Connolly, P., & York, P. (2002). Evaluating capacity-building efforts for nonprofit organizations. *Organizational Development Practitioner, 34*, 33–39.
- Cousins, B., Goh, S., Clark, S., & Lee, L. (2004). Integrating evaluation inquiry into the organizational culture: A review and synthesis of the knowledge base. *Canadian Journal of Program Evaluation, 19*, 99–141.
- Cousins, J. B., Goh, S., & Elliott, C. (2007, Nov.). *Integrating consequences of evaluation into evaluation capacity building inquiry*. Paper presented at the annual meeting of the American Evaluation Association, Baltimore, MD.
- Cousins, J. B., Goh, S. C., Elliott, C., & Aubry, T. (May, 2008). *Government and voluntary sector differences in organizational capacity to do and use evaluation*. Presented at the annual meeting of the Canadian Evaluation Society, Québec. Retrieved from <http://evaluationcanada.ca/site.cgi?s=1>
- Dansec, E., Halsall, T., & Kasprzak, S. (2009). *Readiness assessment tool for evaluation capacity building*. The Provincial Centre of Excellence for Child and Youth Mental Health at CHEO, Ottawa, Canada.
- Duffy, J., Labin, S., & Wandersman, A. (November, 2007). A review of research on evaluation capacity building strategies. In (T., Taylor-Ritzler, Chair) *Building and assessing capacity for evaluation: Creating communities of learners among service providers*. Symposium presented at the annual meeting of the American Evaluation Association Conference, Baltimore, MA.
- Garcia-Iriarte, E., Suarez-Balcazar, Y., & Taylor-Ritzler, T. (2011). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation, 32*, 168–182.
- Gibbs, D., Napa, D., Jolly, D., Westover, B., & Uhl, G. (2002). Increasing evaluation capacity within community-based HIV prevention programs. *Evaluation and Program Planning, 25*, 261–269.
- Gorman-Smith, D., Tolan, P. H., Zelli, A., & Huesmann, L. R. (1996). The relation of family functioning to violence among inner-city minority youths. *Journal of Family Psychology, 10*, 115–129.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation, 29*, 358–368. doi:10.1177/1098214008318798
- Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage.
- King, J. A. (2002). Building the evaluation capacity of a school district. *New Directions for Evaluation, 93*, 63–79. Retrieved from <http://www3.interscience.wiley.com/cgi-bin/fulltext/92013254/PDFSTART>
- King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter, 35*, 10–16.

- Labin, S., Duffy, J., Meyers, D., & Wandersman, A. (2010, November). *Synthesis of evaluation capacity building literature*. Paper presented at the American Evaluation Association Meeting, San Antonio, TX.
- Labin, S., Duffy, J., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation, 33*, 307–338.
- Mackay, K. (2002). The World Bank's ECB experience. *New Directions for Evaluation, 93*, 81–100. Retrieved from <http://www3.interscience.wiley.com/cgi-bin/fulltext/92013252/PDFSTART>
- Milstein, B., Chapel, T. J., Wetterhall, S. F., & Cotton, D. A. (2002). Building capacity for program evaluation at the Centers for Disease Control and Prevention. *New Directions for Evaluation, 93*, 27–46.
- Milstein, B., & Cotton, D. (2000). *Defining concepts for the presidential strand on building evaluation capacity*. Retrieved December 2, 2007, from <http://www.eval.org/eval2000/public/presstrand.pdf>
- Muthén, L. K., & Muthén, B. O. (2009). *Mplus user's guide* (5th ed.). Los Angeles, CA: Author.
- Naccarella, J., Pirkis, J., Kohn, F., Morley, B., Burgess, P., & Blashki, G. (2007). Building evaluation capacity: Definitional and practical implications from an Australian case study. *Evaluation and Program Planning, 30*, 231–236.
- Nielsen, S. B., Lemire, S., & Skov, M. (2011). Measuring evaluation capacity: Results and implications of a Danish study. *American Journal of Evaluation, 32*, 324–344.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation, 29*, 443–459. doi:10.1177/1098214008324182
- Preskill, H., & Torres, R. T. (2000). *Readiness for organizational learning and evaluation instrument*. Available from H. Preskill, Hallie.preskill@fsg.org
- Stevenson, J. F., Florin, P., Mills, D. S., & Andrade, M. (2002). Building evaluation capacity in human service organizations: A case study. *Evaluation and Program Planning, 25*, 233–243.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation, 93*, 7–25.
- Suarez-Balcazar, Y., Taylor-Ritzler, T., Garcia-Iriarte, E., Keys, C. B., Kinney, L., & Ruch-Ross, H., . . . Curtin, G. (2010). Evaluation capacity building: A culturally- and contextually-grounded interactive framework and exemplar. In F. Balcazar, Y. Suarez-Balcazar, T. Taylor-Ritzler, & C. B. Keys (Eds.), *Race, culture and disability: Rehabilitation science and practice* (pp. 307–324). Sudbury, MA: Jones and Bartlett.
- Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation, 28*, 45–59. doi:10.1177/1098214006296430
- Taylor-Ritzler, T., Suarez-Balcazar, Y., & Garcia-Iriarte, E. (November, 2009). Evaluation capacity: Model validation using quantitative results. In (Y., Suarez-Balcazar, Chair), *Evaluation capacity building: Taking stock of and advancing frameworks, strategies, outcomes and measurement*. Presented at the annual meeting of the American Evaluation Association Conference, Orlando, FL.
- Texas Christian University (2005). *Organizational Readiness for Change* (TCU ORC). Institute of Behavioral Research, Forth Worth, TX. Retrieved March 9, 2011, from <http://www.ibr.tcu.edu/evidence/evi-orc.html>
- Tolan, P. H., Gorman-Smith, D., & Henry, D. B. (2003). The developmental-ecology of urban males' youth violence. *Developmental Psychology, 39*, 274–291.
- Volkov, B., & King, J. (2007). *A checklist for building organizational evaluation capacity*. Retrieved October 27, 2008, from <http://www.wmich.edu/evalctr/checklists/ecb.pdf>