WILEY

RESEARCH ARTICLE

Conducting and utilizing evaluation for multiple accountabilities: A study of nonprofit evaluation capacities

Tara Kolar Bryan[1] | Robbie Waters Robichau[2] |
Gabrielle E. L'Esperance[1]

[1]School of Public Administration, University of Nebraska at Omaha, Omaha, Nebraska

[2]Bush School of Government and Public Service, Texas A&M University, College Station, Texas

Correspondence
Tara Kolar Bryan, School of Public Administration, University of Nebraska at Omaha, Omaha, NE.
Email: tbryan@unomaha.edu

Abstract

This article combines the evaluation capacity building and nonprofit accountability literature to examine how nonprofits use evaluation data to address stakeholder expectations. Our study investigates how staff competency, technological resources, learning climate, and strategic planning influence a nonprofit's ability to demonstrate upward, lateral, and downward accountabilities. Results indicate different combinations of evaluation capacities matter more for particular stakeholder groups. We argue a more integrative nonprofit accountability requires that managers and staff know how to *utilize* evaluation results for internal and external audiences. Nuances between specific evaluative capacities and their influences on multiple accountabilities suggest several implications for practice.

KEYWORDS

capacity, evaluation, learning climate, nonprofit accountability, strategic planning

1 | INTRODUCTION

Nonprofit leaders face considerable pressures to demonstrate effectiveness, performance, and efficiency. A growing literature discusses how the "accountability movement" informs and drives nonprofit behavior toward internal and external stakeholders' demands (Carman, 2010; Chaudhary, Diaz, Jayaratne, & Assan, 2020; Costa & Pesci, 2016; Ebrahim, 2005; Williams & Taylor, 2013). The sector's legitimacy depends on nonprofits being accountable to someone for

their performance. And while there are increasing numbers of tools and reporting mechanisms available to show accountability, limited evidence indicates the sector is not vastly improving upon its performance (Ebrahim, 2019, p. 9). Ultimately, accountability entails "being answerable to stakeholders for the actions of the organization, whether by internal or external initiation" (Christensen & Ebrahim, 2006, p. 196). Broad stakeholder interests must be considered by nonprofit leaders. The literature on nonprofit accountability identifies three encompassing stakeholder groups: upward stakeholders, including funders and regulators; lateral stakeholders including staff and volunteers; and downward stakeholders, including participants and others affected by the organization (Chen, Dyball, & Harrison, 2020; Christensen & Ebrahim, 2006; Najam, 1996).

Program evaluation is one of the dominant tools employed to demonstrate accountability (Chen et al., 2020; Hoefer, 2000). An evaluation approach that emphasizes the intended use and users of evaluation data are Utilization-Focused Evaluation (UFE). The central premise of UFE is that evaluation should inform and improve decision-making while also focusing on the needs of those stakeholders who are most likely to care about and use evaluation (Patton, 2012). UFE informs how the intended audiences of nonprofit evaluation data are likely to manifest in three user groups. Upward stakeholders, who desire to fund proven social initiatives, commonly require nonprofit recipients to conduct the evaluation. Funders or regulators may even demand nonprofit staff to participate in trainings to learn the technicalities of conducting evaluation (Connolly & York, 2003). In some cases, these trainings offer limited insights into how to use evaluation results as a tool for improving organizational decision-making (i.e., lateral accountability) and for improving programs and services (i.e., downward accountability) (Leviton, 2014). Emphasizing upward stakeholders' demands over other stakeholders' needs is highly problematic. Specifically, developing evidence indicates program evaluation is not producing intended results for organizations and their clients (Carman, 2010; Hoefer, 2000). In other words, program evaluation aimed at demonstrating the ability to *do* an evaluation for the sake of compliance is not the same as *using* evaluation as a management tool to make informed decisions.

Nonprofit leaders and staff striving to conduct and utilize evaluation may lack the capacities necessary for employing evaluation results in decision-making, limiting their ability to demonstrate accountability (Despard, 2016; Gagnon, Aubry, Cousins, Goh, & Elliott, 2018; Kovner, 2014; Kovner, Elton, & Billings, 2000). The field expects nonprofits to be inclusive of multiple stakeholders' needs, but completing an evaluation does not transform results into a tool for learning and program improvement. A more integrative nonprofit accountability requires that managers and staff know how to *utilize* evaluation results for internal and external audiences. We ask: what organizational capacities influence the likelihood of nonprofit leaders and staff to conduct and use evaluation results to demonstrate accountability to different stakeholder groups?

To that end, this article examines the relationship of four evaluation capacities to upward, lateral, and downward accountabilities. Per the evaluation literature, we frame evaluation capacities in terms of those needed to *do* an evaluation, including technical resources and staff competency, versus capacities for *utilizing* evaluation results as a decision-making tool, including learning climate and strategic planning (Bourgeois, Whynot, & Thériault, 2015; Gagnon et al., 2018; Rogers, Kelly, & McCoy, 2019). We examine these relationships through a Partial Least Squares Structural Equation Model using survey responses from 243 nonprofit executive directors or evaluation staff. Our results indicate that certain evaluation capacities matter more for particular accountability relationships. Our model shows an evaluation *use* variable, for

example, learning climate or strategic planning, is needed for accountability to upward, lateral and downward stakeholders. We conclude that for nonprofits to be accountable, managers and staff must know how to use evaluation findings as a decision-making tool.

1.1 | Multiple nonprofit accountabilities

Determining what nonprofits are accountable for is guided by to whom nonprofits are accountable (Chen et al., 2020; Clark, 1991; Ebrahim, 2016, 2019; Najam, 1996). The term "social accountability" in this context encompasses efforts by nonprofits to maintain accountability to all direct and indirect stakeholders of their organization's work (Costa & Goulart da Silva, 2019). The literature converges around three general stakeholder groups: upward, lateral, and downward. These groups help us understand the breadth of accountability relationships in the nonprofit sector.

Upward accountability is directed toward stakeholders who influence the ability of an organization to pursue its mission such as foundations, donors, watchdog groups, accrediting bodies, or government (Christensen & Ebrahim, 2006; Najam, 1996). Significant attention is given to upward accountability in the literature because nonprofits are most responsive to funder requirements since they provide the critical resources needed to pursue organizational missions. However, the pressure to be accountable upstream directs a nonprofit's focus on meeting funder expectations of operational efficiency, and, on short-term outputs over long-term mission achievement, or, what Ebrahim (2005) calls the "accountability myopia." Accountability myopia causes some nonprofits to overemphasize measurable metrics that are pleasing to funders at the expense of harder to measure long-range goals associated with social change. Increasingly, program evaluation and outcome measurement are required of nonprofit grantees to ensure funders that program outcomes are being met (Ebrahim, 2019; LeRoux, 2009; MacIndoe & Barman, 2012; Yang & Northcott, 2019). Carman (2011) finds the most common motivation for nonprofits to engage in evaluation activities are externally driven. These nonprofit leaders often conduct evaluation chiefly to meet funder requirements and to achieve legitimacy in the eyes of others.

Lateral is the second accountability category, referring to the stakeholders who are accountable to the organization itself, such as staff, board, and volunteers (Christensen & Ebrahim, 2006). Other scholars have referred to this as internal accountability to staff and mission (Van Zyl, Claeyé, & Flambard, 2019). Lateral accountability emphasizes that internal stakeholders should keep the nonprofit's ultimate goal of mission achievement at the forefront of organizational decision processes. The underlying assumption of lateral accountability is that organizations should have processes in place to facilitate learning and data-informed choices. As nonprofit leaders seek greater lateral accountability, evaluation use may lead to more informed management. For example, Hoefer (2000) finds that program evaluations are an important accountability tool for human service nonprofits when they are conducted properly and rigorously, are used by staff to make needed changes, and results are shared with key stakeholders. But these expectations are not always met. Carman (2011) notes that using evaluation as a "rational and technical tool" to inform programmatic choices "did not resonate clearly with most nonprofit executives." (p. 367). This finding illustrates a disconnect between conducting program evaluations and utilizing results for decision-making and stronger lateral accountability.

The third category, downward accountability, is directed toward stakeholders who are affected by the work of the organization such as beneficiaries, members, and the community (Christensen & Ebrahim, 2006). Downward accountability requires nonprofits to offer quality services and program implementation to clients and their beneficiaries who are most affected by the nonprofit's work. The downward accountability literature is limited but burgeoning. Scholars have focused on how nonprofits should be accountable and responsive to beneficiary needs as an imperative for existence (Benjamin, 2012; Ebrahim, 2019; Unerman & O'Dwyer, 2010). Nonprofit leaders need an open and participatory dialogue with participants to achieve broader notions of accountability and program significance (Kraeger & Robichau, 2017; Ospina, Diaz, & O'Sullivan, 2002). Program evaluation and outcome measurement can enhance downward accountability by providing valuable information about how a nonprofit's work ultimately influences their beneficiaries (Benjamin, 2012; Yang & Northcott, 2019).

While the value of program evaluation as an assessment tool is evident, the organizational capacities required to put evaluation findings to use for accountability purposes is not clear-cut. Are there different capacity factors that matter more or less for certain types of accountability? This study addresses this question by assessing the extent to which evaluation capacity factors influence a nonprofit's ability to demonstrate upward, lateral, and downward accountability.

2 | NONPROFIT CAPACITY FOR EVALUATION

The literature on capacity in the nonprofit and public management fields is well established (Andrews & Boyne, 2010; Christensen & Gazley, 2008; Despard, 2017; Honadle, 1981; Letts, Ryan, & Grossman, 1999; Yu-Lee, 2002). Capacity is often conceptualized as the means by which nonprofit organizations achieve their missions. As Shumate, Cooper, Pilny, and Pena-y-lillo (2017) define it, nonprofit capacity incorporates the "processes, practices, and people that the organization has at its disposal that enable it to produce, perform, or deploy resources to achieve its mission" (p. 156). As such, scholars now emphasize the multidimensional nature of organizational capacity as a construct (Bryan, 2019; Christensen & Gazley, 2008; Despard, 2017; Murphy & Robichau, 2016; Shumate et al., 2017). In other words, there is no one nonprofit capacity, rather there are multiple nonprofit *capacities*. Organizational capacity is context-dependent, contingent on the external environment (Christensen & Gazley, 2008) and on how the organization defines effectiveness in a given context (Bryan, 2019). Building on this line of inquiry, we examine organizational capacity in a particular context: the capacity to conduct and utilize evaluation findings to demonstrate accountability to upward, lateral, and downward stakeholders.

Interest in evaluation is growing. External pressures to evaluate programs are numerous: funding desires or requirements (Compton, Baizerman, Preskill, Rieker, & Miner, 2001), direction from the board or other leadership (Taut, 2007), and need for legitimacy (Ebrahim, 2005; Eckerd & Moulton, 2011). Simultaneously, internal motivators, such as the use of performance metrics in strategic decision-making (LeRoux & Wright, 2010) and the adoption of a learning culture are both positively associated with the mission and financial performance (McHargue, 2003). The real challenge lies in harnessing these motivations into a useful management instrument that improves organizations and their programming.

2.1 | Evaluative capacity building

One approach gaining traction among program evaluators is the Evaluative Capacity Building (ECB) framework, which emphasizes the practical applications of evaluation. The goals of the ECB method are using findings to improve organizational outcomes and creating organizational cultures where data utilization becomes an internalized part of daily activities and decision-making. Often, nonprofit leaders that perform evaluation do not use the information they have to influence their programming, in part, because they find the data overwhelming in the face of strategic decision-making (Hoefer, 2000; Snibbe, 2006). To overcome this obstacle, evaluation should be designed for the utility of the data it will produce (Alkin, 1985; Howell & Yemane, 2006; Patton, 2001, 2012), results should be user friendly (Hoefer, 2000), decision-makers should be involved in the educational efforts of the process (Cousins & Leithwood, 1986), and the organization should work toward building its own readiness (Alkin, 1985).

As organizations undergo a change in evaluative capacity, several factors play a role. We highlight four elements of capacity that are particularly germane to evaluation. Two capacity elements relate to an organization's capacity *to do* evaluation: staff competencies in the evaluation and technical resources devoted to evaluation activities. Two capacity elements relate to an organization's capacity *to use* evaluation: learning climate and strategic planning. We hypothesize how each of the evaluation capacities influences upward, lateral, and downward accountability.

2.1.1 | Staff competency in evaluation

The lack of knowledge, skills, and abilities of staff to perform evaluation is an impediment to evaluation practice in nonprofit organizations (Carman, 2009; Carman & Fredericks, 2010; Easterling, 2000; McCoy, Rose, & Connolly, 2013; Mitchell & Berlan, 2016; Thomson, 2011). Organizations are less likely to use evaluation if the staff does not know how to conduct evaluation activities in the first place. In their "State of Evaluation" report, Morariu, Athanasiades, Pankaj, and Grodzicki (2016) find that 48% of respondents identified limited competencies in evaluation knowledge and skills as a core barrier to evaluation practice. After conducting a systematic literature review of barriers and facilitators to an evaluation in the third sector, Bach-Mortensen, Lange, and Montgomery (2018) find that 36 out of 39 studies identified a lack of evaluation understanding in staff as a barrier to evaluation practice. Additional impediments to evaluation include evaluation design challenges, establishing outcome indicators, and data collection and analysis (Bach-Mortensen et al., 2018). Even as organizations strive to build their evaluation capacity, staff may fail to comprehend how evaluation leads to program improvement (Chaudhary et al., 2020). For example, employees may lack an understanding of the nuances of evaluation whether in regard to the overall concepts associated with evaluation (Hoefer, 2000; Liket, Rey-Garcia, & Maas, 2014; Morariu, Athanasiades, & Emery, 2012) or confusion about the processes and desired outcomes of evaluation (Liket et al., 2014; Plantz, Greenway, & Hendricks, 1997; Snibbe, 2006).

Research suggests rigor in evaluation techniques is internally driven. External stakeholders' mandates, whether upward or downward, do not lead to rigorous evaluation (Mitchell & Berlan, 2016). These findings imply that staff competencies in the evaluation will mostly benefit lateral accountability because it focuses on responding to internal stakeholders. We expect that

evaluation competencies of staff increase the likelihood that their nonprofits will use evaluation findings to demonstrate accountability to lateral stakeholders, but not to upward or downward stakeholders. Thus, we test the following hypotheses:

H1c. Upward accountability has no relationship with staff competency in evaluation.
H1b. Lateral accountability is positively associated with staff competency in evaluation.
H1c. Downward accountability has no relationship with staff competency in evaluation.

2.1.2 | Technological resources

There is broad consensus that nonprofit organizations as a whole experience lack of resources for evaluation (Carman, 2007, 2009; Connolly & York, 2002; Ebrahim, 2003; Edwards & Hulme, 1995; Garcia-Iriarte, Suarez-Balcazar, Taylor-Ritzler, & Luna, 2011; McCoy et al., 2013; Morariu et al., 2012; Thomson, 2011). In particular, resources that enable nonprofit staff to perform evaluation work are relevant because most evaluation activities are internally-driven by staff (Carman & Fredericks, 2010; Mitchell & Berlan, 2016). Providing technical resources that support staff in evaluation efforts is an important element of capacity in this context. Technical resources, such as databases, Excel, and program dashboards, help equip staff to enter data and perform analysis (Wade & Kallemeyn, 2020). However, studies suggest nonprofit employees feel their organizations do not allocate the financial and technical resources needed to evaluate their programs well (Morariu et al., 2012; Thomson, 2011).

Even competent staff may struggle to perform evaluation work when technological resources are unavailable. On the other hand, when technological resources are available, staff competency is enhanced by their access to those resources. We expect technological resources to influence staff competency positively and thus posit:

H2. Staff competency has a positive association with technological resources for evaluation.

2.1.3 | Learning climate

Whereas the first two capacity elements (staff competency in evaluation and technical resources) relate to an organization's capability to conduct an evaluation, the last two capacities relate to how the organization incorporates evaluation data in decision-making. Cultivating an organizational climate that values evaluation and data use is a key component of evaluation capacity building (Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry, & Balcazar, 2013). Taylor-Ritzler et al. (2013, p.192) posit that organizational learning is determined by "leadership, culture, systems and structures, and communication." Several organizational and individual factors must coalesce to create learning climates that favor evaluation and data use. For evaluation to become a salient part of organizational progress, learning structures embedded in the evaluation must be applied regularly. However, many nonprofit leaders feel underprepared to take on this responsibility (Mitchell & Berlan, 2016).

The underlying purpose of developing a learning climate in nonprofits is to cultivate a commitment to evaluation that informs decision-making (McCoy et al., 2013). Recent studies suggest that organizational learning climate is an important indicator of evaluation data use in organizations (Cousins, Goh, Elliott, & Bourgeois, 2014; Mitchell & Berlan, 2018; Umar &

Hassan, 2018). For example, Mitchell and Berlan (2018) find that a learning climate positively influences rigor of evaluation techniques and frequency of performance monitoring. Umar and Hassan (2018) contend managers should embrace a "development-oriented climate" where learning activities are integrated throughout organization life and as such, employees are rewarded for these behaviors.

The focus on learning climate relates to learning organizations more broadly (Senge, 1991). The literature underscores inquisitiveness and an openness to new ideas, knowledge sharing, transfer, and integration as indicators of an organization's learning capacity (Jerez-Gomez, Cespedes-Lorente, & Valle-Cabrera, 2005; Strichman, Bickel, & Marshood, 2008). Cultivating a learning climate is particularly critical in determining the extent to which an organization uses evaluation findings internally to inform decisions (Cousins et al., 2014). We expect the learning environment to have a positive relationship with lateral and downward accountability. Ebrahim (2005) proposes that evaluation is the mediator between accountability and organizational learning, but an overemphasis on accountability may come at the expense of organizational learning. The literature further distinguishes between nonprofit managers who foster an evaluation climate that values organizational learning as an end in itself, and those who only emphasize compliance with funder requirements (Winkler & Fyffe, 2016), so we do not expect a relationship between a learning environment and measures of upward accountability. This leads us to hypothesize that:

> *H3a. Upward accountability has no association with organizational learning climate.*
> *H3b. Lateral accountability has a positive association with organizational learning climate.*
> *H3c. Downward accountability has a positive association with organizational learning climate.*

2.1.4 | Strategic planning

Strategic planning is the intentional planning processes, staff and board undertake to formulate a strategy that aligns their internal capabilities with the dynamics of their external environment. Examining evaluation activities from a strategic perspective offers insights into how organizations are accountable to internal and external stakeholders. Research suggests nonprofit managers who embrace evaluation are more likely to utilize data for decision-making and to review their strategic plan regularly (Carman & Fredericks, 2008). Furthermore, strategic planning is associated with improving organizational performance and goal achievement (George, Walker, & Monster, 2019) suggesting that a strategic orientation can assist managers in reporting results to their stakeholders. The underlying assumption of much of this work is that nonprofit managers who are measuring, analyzing, and assessing evaluation data can determine if their strategic goals are being met.

Moreover, leadership commitment plays an important role in both strategic planning efforts and outcome measurement. Applying strategic choice theory, MacIndoe and Barman (2012) examine leaders' perspectives on implementing outcome measurement in nonprofits. They find "the expectations of internal stakeholders (the board of directors and staff) drive patterns of high program use of outcome measurement, regardless of resource allocation" (MacIndoe & Barman, 2012, p. 731). In turn, nonprofit managers value outcome measurement to the extent they perceive internal actors' views as critical to goal achievement. This research illustrates how

leadership support of the strategic approach may positively influence lateral and downward accountability.

Another stream of research emphasizes how managers use strategic planning and evaluation activities to legitimize their organization (Eckerd & Moulton, 2011; Hwang & Bromley, 2015). Relying upon institutional theory, these scholars highlight how nonprofit leaders implement evaluation, strategic planning, and other management tools to satisfy upward stakeholders' demands, not because they intend to implement the plan fully. Moreover, a funder-mandated evaluation may be more symbolic than useful if the funding organization does not consult with nonprofit staff regarding their desired goals from the evaluation (Carman, 2011; Liket et al., 2014; Mitchell & Berlan, 2016). Overall, this literature implies that strategic planning may positively influence upward accountability as well, albeit for very different reasons.

Ultimately, the literature suggests strategic planning can foster the use of evaluation data to inform organizational decision-making and signify legitimacy to stakeholders. Therefore, we posit:

H4a. Upward accountability has a positive association with strategic planning.
H4b. Lateral accountability has a positive association with demonstrating lateral accountability.
H4c. Downward accountability has a positive association with strategic planning.

3 | METHODS

3.1 | Data and sample

We propose the following research question: to what extent does evaluation capacity influence a nonprofit's ability to conduct and use evaluation findings to demonstrate multiple accountabilities? Answering this question brings the evaluation capacity building and the nonprofit accountability literature together in meaningful ways. Unlike many of the nonprofit accountability studies, our study begins to explore how we can quantify nonprofit accountabilities.

This study employs data from nonprofits in one mid-sized, Midwestern city. Working through a local community foundation's listserv of their annual giving day, we administered an online survey to executive directors of 603 organizations in a three-county metropolitan statistical area. Executive directors were asked to complete the survey or to ask their staff member most knowledgeable about the evaluation to complete the survey. Approximately 59% of the respondents were executive directors. Identifying information about our survey respondents and their nonprofits were not collected due to our agreement of anonymity with the foundation, limiting our ability to test for nonresponse biases in the sample. However, organizations surveyed were designated 501(c) 3 as verified by the foundation. The 86-item survey instrument was available from October to December of 2016 via Survey Monkey. Of the 603 distributed surveys, 243 surveys were completed yielding a response rate of 40%.

The sample is diverse in terms of the nonprofit service area and budget size. Respondents come from nonprofits with a wide range of service areas: 43% human service, 17% arts, education and humanity, 15% education and research, 9% public social benefit, 5% health, 3% environment and animals, 8% other. Religious congregations were excluded. The nonprofits were almost evenly distributed in terms of budget size. The largest group of organizations had budgets over $2 million (29%), followed by nonprofit with $500,000 to $2 million (25%), then

$100,000 to $499,999 (24%), and the smallest group had budgets under $100,000 (22%). The diversity of service area and size gives us confidence we will learn something more general about evaluation capacities and their relationships to multiple accountabilities in the nonprofit sector.

We employ the ECB Survey Instrument developed and validated by Taylor-Ritzler et al. (2013). It draws upon a wealth of qualitative research on evaluation capacity building. The survey covers general questions around individual awareness of, motivation to engage in, and competence in evaluation. There are also questions related to organizational factors regarding evaluation such as leadership, work environment, learning climate, and resources. The last section of the survey asks questions about mainstreaming evaluation as part of one's work, using evaluation findings, and other demographics.

3.2 | Measurement

The indicators described here were asked on a five-point Likert scale of agreement (1 = Strongly disagree, 5 = Strongly agree). Each construct is measured using multiple items and is described along with means, standard deviations, and factor loadings in the Appendix. All validity and reliability checks provide satisfactory results (see Table 1 and Appendix).

3.2.1 | Dependent variables

One contribution of this study is to conceptualize and measure quantitatively different accountability types. The research model entails three types of accountability as dependent variables: upward, lateral, and downward accountability. Upward accountability addresses the extent to which evaluation results are reported to funders and to request additional funding. The lateral accountability variable describes how managers integrate and use evaluation data for decision-making and daily workplace practices. Downward accountability measures how evaluation results are used to assess program quality and implementation as well as improve services or programs.

3.2.2 | Independent variables

Another contribution of this study is to explore the relationship between evaluation capacities and nonprofit accountabilities. Three independent variables are constructed in the model to capture capacity. Technical resources measures staff's access to technology and computerized records, and their ability to produce summary reports. Strategic planning captures whether nonprofits have a strategic plan and the organization's goals aligning with strategic priorities. The learning climate describes how the organizational environment encourages evaluation information to be shared in open forums and whether managers and peers can encourage the use of evaluation findings.

3.2.3 | Mediating variables

Staff competency captures the staff's ability to write evaluation reports, define outcome indicators, and decide what questions to answer in an evaluation. As an endogenous variable, staff

TABLE 1 Results of the PLS model estimations

| | Alpha | CR | AVE | Intercorrelations | | | | | | |
				Technological resources	Staff competency	Strategic planning	Learning climate	Upward accountability	Lateral accountability	Downward accountability
Technological resources	.866	0.864	0.760	(0.872)						
Staff competency	.781	0.780	0.542	0.417	(0.736)					
Strategic planning	.922	0.911	0.838	0.179	0.279	(0.915)				
Learning climate	.777	0.768	0.625	0.31	0.358	0.288	(0.796)			
Upward accountability	.821	0.805	0.675	0.179	0.301	0.334	0.315	(0.825)		
Lateral accountability	.751	0.725	0.574	0.419	0.617	0.336	0.758	0.281	(0.758)	
Downward accountability	.903	0.901	0.752	0.274	0.329	0.389	0.692	0.563	0.584	(0.868)

Note: AVE, average variance extracted; CR, components reliability; alpha, Cronbach's alpha; square root of AVE in parentheses.

competency is constructed to mediate the relationship between technical resources lateral accountability.

3.3 | Partial least squares as an analytical technique

Partial least squares structural equation modeling (PLS-SEM) is a components-based estimation approach that integrates path analysis and factor analysis. Like multiple regression analysis, PLS maximizes the explained variance of latent constructs. PLS offers the additional benefit of evaluating data quality through measurement model characteristics and providing robust estimates of the structural model, particularly when sample sizes are smaller and not normally distributed (Chin, 2010; Hair, Ringle, & Sarstedt, 2011). When research goals are theory development and predicting key constructs, PLS-SEM is the preferred method for complex structural models (Hair et al., 2011, pp. 143–144). SmartPLS 3.0 software simultaneously tests theoretical relationships among latent variables (i.e., structural paths) and the relationships between latent variables and their indicators (i.e., measurement paths) (Ringle, Wende, & Will, 2005). The measurement model (Table 1) and the structural model (Table 2) are reported below.

TABLE 2 Hypothesis testing: bootstrapping direct effect results

Path	Hypothesized direction	ß	Estimate	t-value	Decision
H1a: *Staff competency→ upward accountability*	0	.167	0.104	1.61	Supported
H1b: *Staff competency→ lateral accountability*	+	.419	0.087	4.76***	Supported
H1c: *Staff competency→ downward accountability*	0	.052	0.083	0.61	Supported
H2: *Technological resources→ staff competency*	+	.415	0.089	4.70***	Supported
H3a: *Learning climate→ upward accountability*	0	.200	0.104	1.88	Supported
H3b: *Learning climate→ lateral accountability*	+	.505	0.092	5.45***	Supported
H3c: *Learning climate→ downward accountability*	+	.628	0.090	6.90***	Supported
H4a: *Strategic planning→ upward accountability*	+	.231	0.091	2.55**	Supported
H4b: *Strategic planning→ lateral accountability*	+	.075	0.092	0.81	Not supported
H4c: *Strategic planning→ downward accountability*	+	.192	0.088	2.23**	Supported

Note: R^2 for Staff Competency = 0.173; Downward Accountability = 0.527; Lateral Accountability = 0.622; Upward Accountability = 0.192.
**$p < .05$; ***$p < .01$.

4 | RESULTS

Table 1 reports the results of the PLS model estimations, demonstrating reliability as well as convergent and discriminant validity of the item measurements. All Components Reliability (CR) and Cronbach's Alpha (α) are greater than .7 suggesting internal consistency and reliability of the model (for rules of thumb for model evaluation see Hair et al., 2011; also Chin, 2010). The loadings of all items on their factor scores are between .642 and .984 suggesting indicator reliability (Appendix). Items were excluded from the model if the indicator fell below 0.6 rendering a total of 16 items measuring the seven factors. All average variance extracted (AVE) is greater than 0.5 indicating convergence reliability. Each construct meets Fornell and Larcker (1981) criterion for determining discriminant validity, which occurs when the square root of the AVE for each construct is larger than the other constructs in the model.

Table 2 shows the results of structural parameter estimates and hypothesis testing. The PLS structural model is assessed by βs of path coefficients, t-statistics, and R^2. A nonparametric bootstrapping procedure (1,000 replications), incorporated in SmartPLS, is used to evaluate the statistical significance of each path coefficient and to provide confidence that the results were not sample-specific (see Davison & Hinkley, 1997). In this model, technological resources account for 17% of staff competency ($R^2 = 17.3$). The model explains 19% of the variance in upward accountability ($R^2 = 19.2$), 53% in downward accountability ($R^2 = 52.7$); and 62% in lateral accountability ($R^2 = 62.2$). Path coefficients (βs) assess the impacts of the capacity dimensions on upward, lateral, and downward accountability (see Figure 1). All path coefficients are positive but not all of them are significant. Learning climate makes the strongest contribution to downward (β = .628) and lateral (β = .505) accountability followed by staff competency to lateral (β = .419) accountability. In addition, technological resources contribute to staff competency (β = .415).

Table 2 shows the results of the hypotheses. The first set of hypotheses examine the link between staff competency and multiple forms of accountability. The data support the
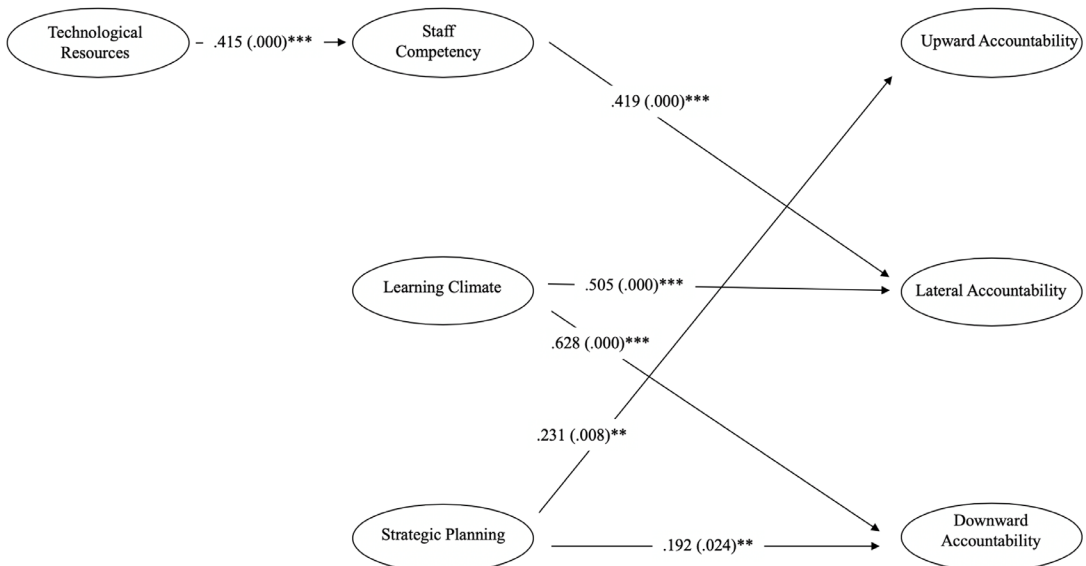

FIGURE 1 Construct model

hypothesis that staff competency is not associated with upward and downward accountability. Hypotheses 1a and 1c, while positive, do not reach the $p < .05$ significance level. However, staff competency is positively related to lateral accountability as suggested by hypotheses 1b ($\beta = .419$, $p < .01$). As suspected, greater staff evaluation competency encourages nonprofits to be more laterally accountable but does not have a significant relationship with upward or downward accountability. Technological resources do have a positive association with staff competency as postulated in hypothesis 2 ($\beta = .415$, $p < .01$). The analysis shows that technological resources indirectly influence lateral accountability through a mediating relationship with staff competency. This means that as access to technological resources grows, staff competency increases, and as this occurs, lateral accountability also increases.

The third set of hypotheses examine the relationship between learning climate and multiple accountabilities. The data show no relationship between learning climate and upward accountability. This finding supports Hypotheses 3a ($p < .104$). The results do confirm a positive association between learning climate and lateral and downward accountability. Hypothesis 3b regarding lateral ($\beta = .505$, $p < .01$) and 3c regarding downward ($\beta = .628$, $p < .01$) accountability state that as learning climate increases so too does lateral and downward accountability. Hypotheses 4a–c focus on the relationships between strategic planning and accountability (We appreciate all the reviewers who helped improve this manuscript. In particular, one reviewer's suggestion to further explore the relationships between strategic planning, learning climate, and the three types of accountability offered additional implications of our paper. We also thank Jooho Lee for his valuable comments on previous drafts of this paper). The data support a positive association between strategic planning and upward and downward accountability. Increases in strategic planning leads to more upward accountability as predicted in hypothesis 4a ($\beta = .231$, $p < .05$) and greater downward accountability as predicted in 4c ($\beta = .192$, $p < .05$). The results do not support a positive relationship between strategic planning and lateral accountability so hypothesis 4b is not confirmed. This means strategic planning is an important factor in increasing upward and downward accountability, but not necessarily lateral accountability.

To further explain these findings, we examined the extent to which strategic planning is directly associated with learning climate, and, whether learning climate acts as a mediator between strategic planning and accountability.[2] The results show that strategic planning does have a direct relationship to learning climate ($\beta = .247$, $p < .000$). The data suggest that strategic planning improves the learning climate of these nonprofits. Furthermore, when a learning climate is modeled as a mediator the relationship between strategic planning and lateral and downward becomes indirect. Learning climate significantly, and positively, mediates the association between strategic planning and lateral accountability ($\beta = .148$, $p < .010$) and downward accountability ($\beta = .179$, $p < .002$). This means that as strategic planning alignment improves, learning climate increases, and as this occurs, lateral and downward accountability also increases. Learning climate, however, does not significantly mediate the relationship between strategic planning and upward accountability ($\beta = .058$, $p < .102$). One explanation for a lack of significance is that while strategic planning signals to funders the existence of a strategic plan, this plan may not be utilized to encourage a learning environment that enhances upward accountability.

A few limitations should be considered before discussing the implications of our findings. Our survey responses rely upon self-reported data. Our sample is based on a subset of nonprofits in one region and may not be generalizable across all regions. While it would

have been helpful, we do not have data on these nonprofit's spending on program evaluation. Considering these limitations, we explore what the data suggest for both research and practice.

5 | DISCUSSION

Our paper contributes to the literature in several meaningful ways. This study combines two literatures, evaluation capacity building with nonprofit accountability, offering insights regarding how nonprofits use evaluation to address stakeholder expectations. The results confirm previous findings and offer new areas for research. Finally, our findings point to nuances between specific evaluative capacities and their influences on multiple accountabilities suggesting several implications for practice.

We examine four particular evaluation capacities: two measuring ability to do the evaluation, staff competencies, and technological resources; and, two measuring the ability to use evaluation findings, learning climate, and strategic planning. The different combinations of capacities matter for each form of accountability. Significantly, the capacity factors associated with the *use* of evaluation findings are present in all three levels of accountability. The capacity factors associated with the ability to *do* evaluation is only significant for lateral accountability. Implications along with future research questions are discussed below.

5.1 | Upward accountability

One core argument of this article is that program evaluation for compliance purposes or funder demands is not synonymous with using evaluation findings for data-informed decision-making. The results at the upward accountability level support this assertion. Our findings underscore a lack of association between the internal capacity to do and use evaluation to demonstrate upward accountability. One explanation is that to answer to funders, a nonprofit will do what it takes to be compliant and survive even in the face of an evaluation capacity deficit. Emphasizing a resource-seeking orientation lends itself to focusing on metrics and evaluations with limited meaning or use in regard to creating social changes and affecting clients and their community (Ebrahim, 2019; Lu Knutsen & Brower, 2010); therefore, having more resources and staff competency in the evaluation may not be as strong of a predictor for using data for upward accountability.

When the evaluation is required by funders, managers may focus on completion rather than using evaluation to encourage a learning climate. By focusing on the requirements for completing the evaluation, choosing external evaluators can be an efficient option. However, problems may arise when evaluation findings are not considered by leaders for more informed decision-making and organizational learning. The lack of findings for upward accountability hints at the ultimate purposes for which funders use evaluation. In many situations, funders believe they are fostering a learning climate by requiring nonprofits to complete program evaluation; however, our findings suggest that having a learning climate is not associated with requiring nonprofits to report evaluation results upward. Our results imply that a different strategy is warranted. If funders want to foster organizational learning climates, they should reward nonprofits that have continuous learning processes and can demonstrate how their learning has positively influenced programs and practices (see McCoy et al., 2013; Umar & Hassan, 2018).

Strategic planning is the only capacity factor positively associated with upward accountability. Effective strategic planning requires organizations to be cognizant of those actors in the external environment that directly affect the organization's ability to pursue missions, such as donors, funding entities, and accrediting agencies. As a symbol of legitimacy, a strategic plan signals to upward stakeholders that the organization is adopting processes that are strategic and forward thinking. What remains unclear, however, is the direction of this relationship. Does the presence of strategic planning processes lead to using evaluation data to demonstrate accountability or vice versa? Further inquiry examining the relationship between evaluation data use for accountability purposes and strategic planning processes is warranted.

5.2 | Lateral accountability

Lateral accountability emphasizes the more internal organizational processes (Van Zyl et al., 2019) that enable data-informed decisions. The results of the lateral accountability model find a significant, positive, and direct association between staff competency and learning climate to lateral accountability. In addition, our results demonstrate that access to technological resources is positive but indirect. Access to technological resources strengthen staff competencies in evaluation, and in turn, staff competencies influence lateral accountability. This makes sense given resources are the inputs that allow nonprofits to conduct an evaluation, and by extension, to use those resources for accountability purposes. Overall, lateral accountability is the only level of accountability that includes both evaluation capacities to do an evaluation (staff competencies) and to utilize evaluation (learning climate).

Those organizations that practice lateral accountability use evaluation data as a management tool. For workers in these organizations, data-informed decision-making is the way things are done. Organizational leaders provide time for staff to engage in evaluation activities daily. In other words, data and learning activities are embedded in daily work practices. Evaluation then becomes a means of learning, not an end in itself. Nonprofits where managers emphasize internal accountability practices value inquiry and application of learning in a substantive way, going beyond program evaluation to integrate data-informed choices at all levels of the organization. Whereas the measures for upward accountability and downward accountability are focused more narrowly on reporting to a funder or assessing quality and implementation of a program, the lateral accountability measures are more expansive incorporating evaluation and data-informed decisions throughout the organization. With this broader understanding of internal accountability in mind, our results suggest both the ability to do and use evaluation is essential.

Lateral accountability challenges nonprofits because it requires leaders simultaneously to embed a culture of learning with the technical capacity to use evaluation to facilitate data-driven choices. The lateral accountability results have similarities to the literature on adaptive capacity. As Connolly and York (2003, p. 2) define it, adaptive capacity is the "ability of a nonprofit organization to monitor, assess, and respond to internal and external changes." Connolly and York and others (see Strichman et al., 2008) argue that adaptive capacity is the most important and complex capacity dimension. It requires technical capacities in tandem with an organizational culture that encourages and promotes new knowledge integration. Future research should explore the link between lateral accountability and adaptive capacity better to understand specific practices nonprofits can adopt to use results that improve organizational

processes internally. More studies are also needed as to how learning climates influence staff competency and their willingness to engage in learning activities in a nonprofit context.

The lack of findings regarding strategic planning and lateral accountability raises further questions over how nonprofits can better integrate strategic goals and priorities as decision-making tools among staff. Without staff buy-in and continual implementation of the strategic plan, the document becomes a relic of leadership rather than a dynamic plan for future action (Bryson, 2018). The lack of finding for lateral accountability echoes a persistent question facing strategic planning and management. Is strategic planning a rational tool of management that informs strategic choices or merely a symbolic signal of organizational legitimacy to external stakeholders? To answer this question, our additional analysis finds that when a strategic plan is incorporated in a larger learning climate, it enhances an organization's ability to use evaluation data for organizational and programmatic decision-making. These results underscore the importance of cultivating learning climates in nonprofit organizations, given its direct relationship with strategic planning, lateral, and downward accountability. These findings also suggest that funding initiatives that support the development of learning climates will likely be money well spent by generating greater levels of both nonprofit accountability and strategic action.

5.3 | Downward accountability

Downward accountability prioritizes beneficiaries and clients by ensuring quality services and program implementation. The results show significant, positive, and direct associations between learning climate and strategic planning with downward accountability. Notably, the downward model includes the capacity factors associated with evaluation data use, not capacity factors that facilitate an organization's ability to conduct an evaluation. In short, neither technical resources nor staff competency matter for demonstrating downward accountability in this study.

While this may seem counterintuitive, the evaluation literacy literature suggests the ability to use and apply evaluation findings is separate from the ability to do an evaluation (Bourgeois et al., 2015; Rogers et al., 2019). In other words, "capacity to use does not first require capacity to do" (Bourgeois et al., 2015, p.47). These scholars argue that what nonprofits need most is understanding and application of evaluation results to organizational practices, not technical competency in designing and conducting the evaluation. Our results indicate that the ability to do evaluation is not required for downward accountability purposes. This finding suggests staff training should prioritize how data use can improve programs and client well-being. Training staff in technical competencies of evaluation may not be as wise of an investment, if those trainings do not include information on how to analyze, interpret, and apply data to programmatic decision-making. While using outside evaluators is common for nonprofits (Carman, 2011; Connolly & York, 2002; Fine, Thayer, & Coghlan, 2003), staff and manager input into the meaning of evaluation results is crucial. These internal stakeholders' ability to interpret and apply evaluation findings to their nonprofit context can further mission achievement.

We also find that strategic planning is positively associated with the evaluation use for downward accountability. Effective strategic planning requires organizations to be inclusive of all stakeholder groups (e.g., board, beneficiaries, staff, & funders) throughout the process; therefore, it offers a plan for not only whose voice is heard, but also to whom the organization is responsible for its results. Strategic managers are boundary spanners (Isbell, 2012; Williams, 2002) who may connect evaluation data to external audiences for continuous improvement purposes (Carman, 2011). Moreover, our additional analysis finds that when

learning climate is modeled as a mediator between strategic planning and downward accountability, learning climate significantly and positively influences the relationship between strategic planning and downward accountability. This finding provides compelling support for the development of learning climates in nonprofit organizations that increases accountability while also strengthening an organization's ability to meaningfully implement strategic plans.

6 | CONCLUSION

This study provides evidence that different combinations of evaluation capacities are linked to upward, lateral, and downward accountability in specific, measurable ways. By employing nonprofit managers and staff this research enhances our understanding of the breadth of accountability relationships in nonprofits. The future inquiry should further investigate how managers address the needs of overlapping stakeholders (e.g., a volunteer who also donates) to gain more insights into these accountability relationships. Exploring the intersection between accountability and capacity also raises questions for nonprofit managers whose organizations perform evaluation but have not thought about how to use evaluation results for multiple forms of accountability. While the data are unable to speak to the evaluation quality or approaches themselves (e.g., utilization-focused - or program-oriented; formative vs. summative, etc.), it is probable that rigor and evaluation type shapes nonprofit accountability and vice versa. Correspondingly, nonprofits at different stages of the organizational life cycle—and with varying levels of evaluative capacities—likely prioritize different accountability relationships (e.g., start-up vs. a mature organization).

Our research shows how managers harness their nonprofits' evaluative capacities to *do* and *use* evaluation in ways that influence their ability to demonstrate accountability. We argue that knowing how to utilize evaluation as a management decision-making tool is as important, if not more important than being technically competent in evaluation techniques for accountability purposes. This article began by suggesting program evaluation for compliance purposes is not equal to using evaluation findings to make data-informed decisions. As long as evaluation efforts focus on performance metrics that do not encourage data utilization, managers may struggle to capitalize on the opportunities evaluation results offer to improve performance. Nonprofit managers and staff should value the results of the evaluation for their own learning and decision-making purposes, but they will not do this if they only see it as a tool of upward accountability. In conclusion, those working from either the funder or evaluator perspective must learn to negotiate what forms of accountability matters most for organizational decision-making in the beginning of evaluation design: compliance to funding requirements, organizational learning, program improvements, or client well-being. Rather than having to choose between these options, managers can strive for more integrative, UFEs that meet the needs of all stakeholder groups while leading to better decisions.

ORCID

Tara Kolar Bryan ● https://orcid.org/0000-0002-0961-8515
Robbie Waters Robichau ● https://orcid.org/0000-0002-9038-5463
Gabrielle E. L'Esperance ● https://orcid.org/0000-0003-3231-4306

REFERENCES

Alkin, M. C. (1985). *A guide for evaluation decision makers*. Beverly Hills, CA: SAGE Publications Incorporated.

Andrews, R., & Boyne, G. A. (2010). Capacity, leadership, and organizational performance: Testing the black box model of public management. *Public Administration Review*, *70*(3), 443–454.

Bach-Mortensen, A. M., Lange, B. C., & Montgomery, P. (2018). Barriers and facilitators to implementing evidence-based interventions among third sector organizations: A systematic review. *Implementation Science*, *13*(1), 103.

Benjamin, L. M. (2012). The potential of outcome measurement for strengthening nonprofits' accountability to beneficiaries. *Nonprofit and Voluntary Sector Quarterly*, *42*(6), 1224–1244.

Bourgeois, I., Whynot, J., & Thériault, É. (2015). Application of an organizational evaluation capacity self-assessment instrument to different organizations: Similarities and lessons learned. *Evaluation and Program Planning*, *50*, 47–55.

Bryan, T. K. (2019). Toward a contingency model for the relationship between capacity and effectiveness in nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, *48*(4), 885–897.

Bryson, J. M. (2018). *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement* (5th ed.). Hoboken, NJ: John Wiley & Sons.

Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation*, *28*(1), 60–75.

Carman, J. G. (2009). Nonprofits, funders, and evaluation: Accountability in action. *The American Review of Public Administration*, *39*(4), 374–390.

Carman, J. G. (2010). The accountability movement: What's wrong with this theory of change? *Nonprofit and Voluntary Sector Quarterly*, *39*(2), 256–274.

Carman, J. G. (2011). Understanding evaluation in nonprofit organizations. *Public Performance & Management Review*, *34*(3), 350–377.

Carman, J. G., & Fredericks, K. A. (2008). Nonprofits and evaluation: Empirical evidence from the field. *New Directions for Evaluation*, *2008*(119), 51–71.

Carman, J. G., & Fredericks, K. A. (2010). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *American Journal of Evaluation*, *31*(1), 84–104.

Chaudhary, A. K., Diaz, J., Jayaratne, K. S. U., & Assan, E. (2020). Evaluation capacity building in the nonformal education context: Challenges and strategies. *Evaluation and Program Planning*, *79*, 101768.

Chen, J., Dyball, M. C., & Harrison, G. (2020). Stakeholder salience and accountability mechanisms in not-for-profit service delivery organizations. *Financial Accountability & Management*, *36*(1), 50–72.

Chin, W. W. (2010). How to write up and report PLS analysis. In E. V. Vinzi, W. W. Chin, J. Hensler, & H. W. Wang (Eds.), *Handbook of partial least squares* (pp. 655–690). Heidelberg, Germany: Springer.

Christensen, R. A., & Ebrahim, A. (2006). How does accountability affect mission? The case of a nonprofit serving immigrants and refugees. *Nonprofit Management & Leadership*, *17*(2), 195–209.

Christensen, R. K., & Gazley, B. (2008). Capacity for public administration: Analysis of meaning and measurement. *Public Administration and Development: The International Journal of Management Research and Practice*, *28*(4), 265–279.

Clark, J. (1991). *Democratizing development: The role of voluntary organizations (no. E51 C593 GTZ-PR)*. West Hartford, CT: Kumarian.

Compton, D., Baizerman, M., Preskill, H., Rieker, P., & Miner, K. (2001). Developing evaluation capacity while improving evaluation training in public health: The American Cancer Society's collaborative evaluation fellows project. *Evaluation and Program Planning*, *24*(1), 33–40.

Connolly, P., & York, P. (2002). Evaluating capacity-building efforts for nonprofit organizations. *OD Practitioner*, *34*(4), 33–39.

Connolly, P., & York, P. (2003). *Building the capacity of capacity builders to last: A Grantmaker's guide to strengthening nonprofit organizations*. Philadelphia, PA: The Conservation Company.

Costa, E., & Goulart da Silva, G. (2019). Nonprofit accountability: The viewpoint of the primary stakeholders. *Financial Accountability & Management*, 35(1), 37–54.

Costa, E., & Pesci, C. (2016). Social impact measurement: Why do stakeholders matter? *Sustainability Accounting, Management and Policy Journal*, 7(1), 99–124.

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.

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, 56(3), 331–364.

Davison, A., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

Despard, M. R. (2016). Challenges in implementing evidence-based practices and programs in nonprofit human service organizations. *Journal of Evidence-Informed Social Work*, 13(6), 505–522.

Despard, M. R. (2017). Can nonprofit capacity be measured? *Nonprofit and Voluntary Sector Quarterly*, 46(3), 607–626.

Easterling, D. (2000). Using outcome evaluation to guide grantmaking: Theory, reality, and possibilities. *Nonprofit and Voluntary Sector Quarterly*, 29(3), 482–486.

Ebrahim, A. (2003). Accountability in practice: Mechanisms for NGOs. *World Development*, 31(5), 813–829.

Ebrahim, A. (2005). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly*, 34(1), 56–87.

Ebrahim, A. (2016). The many faces of nonprofit accountability. In R. O. David (Ed.), *Jossey-bass handbook of nonprofit leadership and management* (4th ed., pp. 102–123). Hoboken, NJ: John Wiley & Sons.

Ebrahim, A. (2019). *Measuring social change: Performance and accountability in a complex world*. Stanford: Stanford University Press.

Eckerd, A., & Moulton, S. (2011). Heterogeneous roles and heterogeneous practices: Understanding the adoption and uses of nonprofit performance evaluations. *American Journal of Evaluation*, 32(1), 98–117.

Edwards, M., & Hulme, D. (1995). NGO performance and accountability in the post-cold war world. *Journal of International Development*, 7(6), 849–856.

Fine, A. H., Thayer, C. E., & Coghlan, A. (2003). Program evaluation practice in the nonprofit sector. *Nonprofit Management & Leadership*, 10(3), 331–339.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 19(4), 39–50.

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.

Garcia-Iriarte, E., Suarez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2011). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation*, 32(2), 168–182.

George, B., Walker, R. M., & Monster, J. (2019). Does strategic planning improve organizational performance? A meta-analysis. *Public Administration Review*, 79(6), 810–819.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19(2), 139–152.

Hoefer, R. (2000). Accountability in action?: Program evaluation in nonprofit human service agencies. *Nonprofit Management & Leadership*, 11(2), 167–177.

Honadle, B. W. (1981). A capacity-building framework: A search for concept and purpose. *Public Administration Review*, 41(5), 575–580.

Howell, E. M., & Yemane, A. (2006). An assessment of evaluation designs: Case studies of 12 large federal evaluations. *American Journal of Evaluation*, 27(2), 219–236.

Hwang, H., & Bromley, P. (2015). Internal and external determinants of formal plans in the nonprofit sector. *International Public Management Journal*, 18(4), 568–588.

Isbell, M. G. (2012). The role of boundary spanners as the interorganizational link in nonprofit collaborating. *Management Communication Quarterly*, 26(1), 159–165.

Jerez-Gomez, P., Cespedes-Lorente, J., & Valle-Cabrera, R. (2005). Organizational learning capability: A proposal of measurement. *Journal of Business Research*, 58(6), 715–725.

Kovner, A. R. (2014). Evidence-based management: Implications for nonprofit organizations. *Nonprofit Management & Leadership*, *24*(3), 417–424.

Kovner, A. R., Elton, J. J., & Billings, J. (2000). Evidence-based management. *Frontiers of Health Services Management*, *16*, 3–24.

Kraeger, P., & Robichau, R. (2017). Questioning stakeholder legitimacy: A philanthropic accountability model. *Journal of Health and Human Services Administration*, *39*(4), 470–519.

LeRoux, K. (2009). Managing stakeholder demands: Balancing responsiveness to clients and funding agents in nonprofit social service organizations. *Administration and Society*, *41*(2), 158–184.

LeRoux, K., & Wright, N. S. (2010). Does performance measurement improve strategic decision making? Findings from a national survey of nonprofit social service agencies. *Nonprofit and Voluntary Sector Quarterly*, *39*(4), 571–587.

Letts, C., Ryan, W., & Grossman, A. (1999). *High performance nonprofit organizations: Managing upstream for greater impact (nonprofit law, finance, and management series)*. New York: Wiley.

Leviton, L. C. (2014). Some under examined aspects of evaluation capacity building. *American Journal of Evaluation*, *35*(1), 90–94.

Liket, K. C., Rey-Garcia, M., & Maas, K. E. (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.

Lu Knutsen, W., & Brower, R. S. (2010). Managing expressive and instrumental accountabilities in nonprofit and voluntary organizations: A qualitative investigation. *Nonprofit and Voluntary Sector Quarterly*, *39*(4), 588–610.

MacIndoe, E., & Barman, E. (2012). How organizational stakeholders shape performance measurement in nonprofits: Exploring a multidimensional measure. *Nonprofit and Voluntary Sector Quarterly*, *42*(4), 716–738.

McCoy, A., Rose, D., & Connolly, M. (2013). Developing evaluation cultures in human service organisations. *Evaluation Journal of Australasia*, *13*(1), 15–20.

McHargue, S. K. (2003). Learning for performance in nonprofit organizations. *Advances in Developing Human Resources*, *5*(2), 196–204.

Mitchell, G. E., & Berlan, D. (2016). Evaluation and evaluative rigor in the nonprofit sector. *Nonprofit Management & Leadership*, *27*(2), 237–250.

Mitchell, G. E., & Berlan, D. (2018). Evaluation in nonprofit organizations: An empirical analysis. *Public Performance & Management Review*, *41*(2), 415–437.

Morariu, J., Athanasiades, K., & Emery, A. (2012). *State of evaluation 2012: Evaluation practice and capacity in the nonprofit sector*. Washington, DC: Innovation Network. Retrieved fromhttp://www.innonet./org/client_docs/innonet-state-of-evaluation-2012.pdf.

Morariu, J., Athanasiades, K., Pankaj, V., & Grodzicki, D. (2016). *State of evaluation 2016: Evaluation practice and capacity of the nonprofit sector*. Washington, DC: Innovation Network, Retrieved from. https://www.innonet.org/news-insights/resources/state-of-evaluation-2016-evaluation-capacity-and-practice-in-the-nonprofit-sector/

Murphy, H., & Robichau, R. W. (2016). Governmental influences on organizational capacity: The case of child welfare nonprofits. *Nonprofit Policy Forum*, *7*(3), 339–367.

Najam, A. (1996). Understanding the third sector: Revisiting the prince, the merchant, and the citizen. *Nonprofit Management & Leadership*, *7*(2), 203–219.

Ospina, S., Diaz, W., & O'Sullivan, J. F. (2002). Negotiating accountability: Managerial lessons from identity-based nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, *31*(1), 5–31.

Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high-quality lessons learned. *American Journal of Evaluation*, *22*(3), 329–336.

Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Los Angeles: Sage.

Plantz, M. C., Greenway, M. T., & Hendricks, M. (1997). Outcome measurement: Showing results in the nonprofit sector. *New Directions for Evaluation*, *1997*(75), 15–30.

Ringle, C. M., Wende, S., & Will, A. (2005). *Smartpls 2.0 M3 (Beta)*. Hamburg, University of Hamburg.

Rogers, A., Kelly, L. M., & McCoy, A. (2019). Evaluation literacy: Perspectives of internal evaluators in non-government organizations. *The Canadian Journal of Program Evaluation*, *34*(1), 1–20.

Senge, P. (1991). The fifth discipline, the art and practice of the learning organization. *Performance and Instruction*, *30*(5), 37–37. https://doi.org/10.1002/pfi.4170300510

Shumate, M., Cooper, K. R., Pilny, A., & Pena-y-lillo, M. (2017). The nonprofit capacities instrument. *Nonprofit Management & Leadership*, *28*(2), 155–174.

Snibbe, A. C. (2006). Drowning in data. *Stanford Social Innovation Review*, *4*(3), 39–45.

Strichman, N., Bickel, W. E., & Marshood, F. (2008). Adaptive capacity in Israeli social change nonprofits. *Nonprofit and Voluntary Sector Quarterly*, *37*(2), 224–248.

Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation*, *28*(1), 45–59.

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.

Thomson, D. E. (2011). The role of funders in driving nonprofit performance measurement and use in strategic management. *Public Performance & Management Review*, *35*(1), 54–78.

Umar, S., & Hassan, S. (2018). Encouraging the collection of performance data in nonprofit organizations: The importance of organizational support for learning. *Public Performance & Management Review*, *42*(5), 1062–1084.

Unerman, J., & O'Dwyer, B. (2010). NGO accountability and sustainability issues in the changing global environment. *Public Management Review*, *12*(4), 475–486.

Van Zyl, W. H., Claeyé, F., & Flambard, V. (2019). Money, people or mission? Accountability in local and non-local NGOs. *Third World Quarterly*, *40*(1), 53–73.

Wade, J., & Kallemeyn, L. (2020). Evaluation capacity building (ECB) interventions and the development of sustainable evaluation practice: An exploratory study. *Evaluation and Program Planning*, *79*, 101777.

Williams, A. P., & Taylor, J. A. (2013). Resolving accountability ambiguity in nonprofit organizations. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *24*, 559–580.

Williams, P. (2002). The competent boundary spanner. *Public Administration*, *80*(1), 103–124.

Winkler, M. K., & Fyffe, S. D. (2016). *Strategies for cultivating an organizational learning culture*. Washington, DC: Urban Institute.

Yang, C., & Northcott, D. (2019). How can the public trust charities? The role of performance accountability reporting. *Accounting and Finance*, *59*(3), 1681–1707.

Yu-Lee, R. (2002). *Essentials of capacity management (essentials series)*. New York: Wiley.

AUTHOR BIOGRAPHIES

Tara Kolar Bryan, Ph.D. is an associate professor and chair of the MPA program at the University of Nebraska at Omaha. Her research interests include nonprofit capacity, foundation-funded capacity building programs and strategic management. Her work has appeared in Nonprofit and Voluntary Sector Quarterly, Nonprofit Policy Forum, Human Service Organizations: Management, Leadership, & Governance, among other venues.

Robbie Waters Robichau, Ph.D. is an assistant professor in the Bush School of Government and Public Service at Texas A&M. Her research examines issues of nonprofit management, organizational accountability and capacity, and meaningfulness in public service work. Her work has appeared in American Review of Public Administration, Nonprofit Policy Forum, Human Service Organizations: Management, Leadership, & Governance, Policy Studies Journal, among other venues.

Gabrielle E. L'Esperance is a doctoral candidate in the School of Public Administration at the University of Nebraska at Omaha. Her research focuses on African non-governmental

organization leadership, with an emphasis on the use of organizational agency and everyday creativity in relationships with global funders.

APPENDIX: Survey Items and Statistics

Construct and items	Mean	SD	Factor loadings
Upward accountability (*Please indicate the extent to which your program currently uses evaluation results for the following purposes*)			
To report to a funder	3.65	1.06	0.903*
To get additional funding	3.57	1.04	0.731*
Lateral accountability (*Please indicate the extent to which the following statements relate to your work*)			
I am able to integrate evaluation activities into my daily work practices	3.35	1.03	0.858*
I have access to the information I need to make decisions regarding my work	3.76	.888	0.642*
Downward accountability (*Please indicate the extent to which your program currently uses evaluation results for the following purposes*)			
To assess quality of a program	3.75	.940	0.850*
To improve services or programs	3.77	.980	0.911*
To assess implementation of a program	3.51	.993	0.839*
Staff competency (*I know how to...*)			
Write an evaluation report	3.12	1.02	0.721*
Define outcome indicators of my programs	3.46	1.02	0.726*
Decide what questions to answer in an evaluation	3.22	.930	0.760*
Technological resources (*In my program...*)			
Staff has access to technology to compile information from computerized records	3.45	1.09	0.842*
Staff has access to adequate technology to produce summary reports of information collected from	3.44	1.05	0.901*
Learning climate (*The organization where I work fosters an environment in which...*)			
Evaluation information is shared in open forums	3.74	.959	0.723*
Staff can encourage managers and peers to make use of evaluation findings	3.70	.868	0.852*
Strategic planning (*Please indicate the extent to which the following statements relate to your organization*)			
My organization has a strategic plan, approved by the board	4.12	1.08	0.841*
My organization's goals are tied to strategic priorities in the strategic plan	4.06	1.02	0.984*

Note: All factor loadings are significant at the *$p < .001$ level, $N = 243$.