

Research Article

The Adoption of Outcome Measurement in Human Service Nonprofits

Chongyoung Lee – California State University, Los Angeles

Richard M. Clerkin – North Carolina State University

Performance measurement has received increasing attention in the nonprofit sector. While the current scholarship has emphasized the balanced use of performance measures, only a limited number of nonprofits have adopted outcome measurement. In this study, we investigate what factors explain a variation in the adoption of outcome measurement based on survey data from 263 human service nonprofits. The results of logistic regression find that human service nonprofits are more likely to adopt outcome measures when they are more risk-taking, influenced by institutional pressures, and have higher human resource and political capacities.

Keywords: Outcome Measurement, Performance Evaluation, Nonprofit Organizations

Performance measurement has received growing academic and practitioner attention in the nonprofit sector due to increasing emphasis on accountability in nonprofits' funding and competitive pressures within the sector (Cairns, Harris, Hutchison, & Tricker, 2005; Ebrahim & Rangan, 2010; Lee & Nowell, 2015; Moxham, 2009). Furthermore, according to a survey of public charity and private foundation employees conducted by McLean and Brouwer (2010), about 40% of respondents answered that the total contributions to their organizations decreased in 2010 compared with the previous year. In response to these pressures, performance measurement in the nonprofit sector has been increasingly emphasized as a way to document an organization's impact to funders as they compete for a shrinking pool of resources.

Scholars have developed a variety of performance measurement tools for nonprofit organizations (Lee & Nowell, 2015). One way to measure performance is through a logic model, which is defined as "a tool used to help identify the linkage between program activities and outputs to desired outcomes" (Hatry, Lampkin, Morley, & Cowan, 2003, p. 3). Using a logic model, organizations can conceptualize their performance as inputs, processes, outputs, and outcomes. Inputs refer to the financial, human, organizational, and community resources that go into programs. Processes are the program activities conducted using the inputs. Outputs refer to the products or services directly produced through program activities, which are typically measured by the number of people served, frequency of services provided, number of goods transferred to the clients, and so on. Last, outcomes are the substantial changes, benefits, or values resulting from the programs and services, which are generally measured by the program participants' modified behaviors, improved living conditions, and the increased skills or knowledge of people served (Ebrahim & Rangan, 2010; Lee & Nowell, 2015; Taylor-Powell & Henert, 2008; W.K. Kellogg Foundation, 2004).

In particular, outcome measurement has become important. Bohte and Meier (2000) highlight that the absence of outcome measures deprives organizations of important information about their effectiveness. Lack of systematic and thoughtful data collection can produce goal displacement because nonprofits may focus on generating easily measureable data or figures that please stakeholders, rather than accomplishing the more meaningful, but more difficult to measure outcomes. If nonprofits use only output measures, they may single-mindedly focus on

Lee, C., & Clerkin, R. M. (2017). The adoption of outcome measurement in human service nonprofits. *Journal of Public and Nonprofit Affairs*, 3(2), 111-134. doi:10.20899/jpna.3.2.111-134

achieving outputs, neglecting the longer-term performance measured by outcomes (Bohte & Meier, 2000). This form of goal displacement can result in mission drift, which is defined as “a diversion of time, energy, and money away from a nonprofit’s mission” (Jones, 2007, p. 300). Outcome measurement is considered as a key to accountability in the nonprofit sector (Moxham, 2009).

The pressure to adopt outcome measurement systems is particularly acute for human service nonprofits. Most human service nonprofits receive funding from federal and state governments but function outside a system of government. Accordingly, diverse mechanisms to assure proper accountability of human service nonprofits have received increasing attention (Dicke, 2002; Dicke & Ott, 1999; Kim, 2005; Salamon, Sokolowski, & Sturza, 1999). In particular, human service nonprofits have increasingly adopted outcome measurement because outcome information is required in many grantmaking processes (Carrilio, Packard, & Clapp, 2004; Horsch, 1996; Julian & Clapp, 2000). Human service nonprofits face additional challenges because of complex accountability relationships with multiple stakeholders who have diverse, often conflicting, outcome expectations (Kim, 2005). Thus, more scholarly efforts are required in the outcome measurement of human service nonprofits to address this challenge.

However, most nonprofits prefer output measures to outcome measures (Wiener, Kirsch, & McCormack, 2002). One reason seems to be the assumption that measuring outcomes is too expensive and too slow to provide quick feedback about an organization’s performance (Moore, 2003; Moxham, 2009). In this regard, only a limited number of nonprofits have adopted outcome measurement, and, correspondingly, several researchers have reported a variation in the adoption of outcome measurement (e.g., Barman & MacIndoe, 2012; Carman, 2007; Carman & Fredericks, 2008; Eckerd & Moulton, 2010; Thomson, 2010). To date, however, limited scholarship attention has been paid toward the reasons for this variation. Hence, in this study, we examine what leads nonprofits to adopt outcome measurement, and whether they report collecting data for specific outcome measures relevant to human services nonprofits. How and whether they use this data is beyond the scope of this study.

To help answer this question, we pull from three theory bases—agency theory, institutional theory, and organizational capacity—to explore potential causes for variation in the adoption of outcomes measurement by human service nonprofits. In the balance of this article, we will briefly review agency theory, institutional theory, and studies of organizational capacity to develop testable hypotheses about factors that can influence the adoption of outcome measurement by human service nonprofits. Then, we will explain the methods we use to test the hypotheses, outline key findings, and discuss the results. Finally, we will close with the implications, limitations, and contributions of our research.

Theoretical Development

Agency Theory

Agency theorists have sought to understand organizational behaviors in light of a principal–agent relationship. Their approach is based on the premise that principals and agents place greater value on their own personal gains rather than on collective goals. Correspondingly, goal conflict between principals and agents is assumed. Various incentives and sanctions for resolving the conflict and achieving goal alignment have been discussed, and redistributing (or transferring) risk between principal and agent is one of them (Davis, Schoorman, & Donaldson, 1997; Eisenhardt, 1989; Van Slyke, 2007).

In addition, agency theory postulates that the principal has imperfect information, thus not knowing exactly what the agent is doing or has done and cannot determine if the agent's behaviors are appropriate. Further, it is assumed that the agent uses this information asymmetry for its own interests, which results in opportunism (Benjamin, 2008; Eisenhardt, 1989; Van Slyke, 2007). To deal with this agency problem, the principal has two options. The first is developing an information system to reveal and control the agent's behaviors; the second is contracting on the "outcomes of the agent's behavior." Because the first option is outside the scope of our study, we focus on the second option. Principals can align the agents' interests and preferences with theirs via an outcome-based contract, "but at the price of transferring risk to the agent" (Eisenhardt, 1989, p. 61). Contracting on outcomes means that the agents are responsible for outcomes, which signifies the agents accept the risk. Accordingly, when the agents are more risk averse, they are less willing to use an outcome-based contract (Benjamin, 2008; Eisenhardt, 1989). Further, agents using an outcome-based contract would be extrinsically motivated to accomplish the outcomes in order to maximize its gains (Amaratunga & Baldry, 2002; Tornow, 1993). In brief, agency theory suggests that risk-taking nonprofits are more likely to use outcome based contracts, and these nonprofits are more likely to adopt outcome measurement. It is thus theoretically reasonable to expect a negative relationship between risk aversion and the adoption of outcome measurement.

H₁: Risk aversion of nonprofit organizations is negatively associated with the adoption of outcome measurement.

Institutional Theory

Institutional theory holds that organizations seek to obtain legitimacy because it enhances organizational growth and survival by providing means to acquire various resources such as capital, human resources, technology, and networks (Aldrich & Fiol, 1994; Dart, 2004; Eckerd & Moulton, 2010; Meyer & Rowan, 1977; Zimmerman & Zeitz, 2002; Zucker, 1987). In this regard, Dart (2004) argues that institutional theory is "built around the concept of legitimacy" (p. 415). According to Suchman (1995), legitimacy is defined as "a generalized perception or assumption that the actions of an entity are socially desirable, proper or appropriate within some socially constructed system of norms, value, beliefs and definitions" (p. 574). Literature highlights that organizations can obtain the legitimacy by achieving conformity with institutional isomorphic pressures, which include coercive, mimetic, and normative isomorphism (e.g., Dart, 2004; Frumkin & Galaskiewicz, 2004).

According to DiMaggio and Powell (1983), coercive isomorphism stems from "both formal and informal pressures exerted on organizations by other organizations upon which they are dependent and by cultural expectations in the society within which organizations function" (p. 150). Resource providers of nonprofits may mandate outcome measurement in the grant-making process. For example, many local United Ways require the nonprofits they fund to measure outcomes (Julian & Kombarakaran, 2006; MacIndoe & Barman, 2012). The mandates from higher authorities may provide nonprofits with an impetus for adopting outcome measurement. In consideration of this line of thought, we expect:

H₂: Program funding tied to mandatory outcome measurement is positively related to the adoption of outcome measurement.

In addition to coercive isomorphism, organizations can enhance their legitimacy by conforming to mimetic pressures. Organizations tend to imitate other organizations' behaviors or practices, which are perceived to be successful when technologies are not well-defined, when goals are

unclear, or when their organizational environment is uncertain (DiMaggio & Powell, 1983; Haveman, 1993). When organizations perceive that other organizations become successful by adopting some practices (e.g., performance measurement practices), the perceptions may engender “legitimacy concerns among remaining non-adopters,” which motivates them to imitate the early adopters’ behaviors (Westphal, Gulati, & Shortell, 1997, p. 371). “Even if the change is only *pro forma*” (Bolman & Deal, 1991, p. 278), the imitating behavior makes the organizations look proactive, modern, and responsive, which legitimizes them (Roy & Seguiné, 2000). By following this line of logic, we expect:

H_{3a}: Mimetic pressure from the other organizations in the same field adopting outcomes measurement positively influences an organization’s adoption of outcome measurement.

In addition, risk aversion may influence the relationship between mimetic isomorphism and the adoption of outcome measurement. As previously hypothesized, mimetic pressure may be positively related to the adoption of outcome measurement. Further, when nonprofits are highly risk averse, the probability of imitation may be increased. According to Lieberman and Asaba (2006), mimetic behavior mitigates risk for any given organization “while raising the risk of failure for those that did not follow” (p. 367). Similarly, Knickerbocker (1973) argues that risk minimization leads to “follow the leader” behavior. When organizations adopt similar practices, there is little chance of a certain organization’s success or failure relative to others. Although differentiation may lead to organizational success, as organizations are more risk averse, they would be more likely to want to look like other organizations rather than look different, which can be highly risky (Lieberman & Asaba, 2006). It is thus reasonable to predict:

H_{3b}: Risk aversion has a positive interaction on the relationship between mimetic isomorphic pressure and the adoption of outcome measurement.

As described in the previous paragraph, we expect that mimetic isomorphism is positively related to the adoption of outcome measurement. Additionally, institutional theorists assume that organizations imitate the others’ behaviors or practices in response to uncertainty. To put it another way, they deal with the uncertainty by modeling themselves after other organizations in the same field (DiMaggio & Powell, 1983; Ellis & Shpielberg, 2003). Consistent with this, Mizruchi and Fein (1999) find that organizations choose to mimic a successful organization when they face an uncertain situation, and DiMaggio and Powell (1983) highlight that uncertainty is a force encouraging imitation. In other words, when an organizational environment creates an uncertainty, it may reinforce other organizations’ mimetic behaviors, which may include the adoption of outcome measurement. Therefore, we expect:

H₄: Uncertainty has a positive interaction on the relationship between mimetic isomorphic pressure and the adoption of outcome measurement.

Normative isomorphism results from professionalism, which is defined as “collective struggle of members of an occupation to define the conditions and methods of their work” (DiMaggio & Powell, 1987, p. 152). Nonprofits, which historically are characterized as amateurish, have been strongly affected by expanded professionalization in contemporary society (Hwang & Powell, 2009; Karl, 1998). Nonprofit organizations may learn about the importance of outcome measurement and share its value through the processes of professionalization. Two mechanisms have been discussed in the literature as the main sources of the professionalism. One is the development of organizational norms by universities and other education institutions. A second is diffusion of the norms and values through professional networks or trade associations

(Barman & MacIndoe, 2012; DiMaggio & Powell, 1987; Fligstein, 1985; Hwang & Powell, 2009; Mizruchi & Fein, 1999; Radaelli, 2000). These processes of professionalization may form a normative pressure for measuring outcomes, which motivates nonprofits to adopt outcome measurement. It is thus theoretically reasonable to expect:

H_{5a}: Professionalism is positively associated with the adoption of outcome measurement.

Further, the positive relationship between professionalism and outcome measurement assumed above may depend on the organization's attitude toward risk. Risk-adverse organizations might be less responsive to the normative pressure for outcome measurement resulting from professionalization. Professionalization may involve organizational changes (e.g., filtering of personnel, restructuring) in the process of developing and promulgating organizational norms regarding professional practices and behaviors (DiMaggio & Powell, 1987). As Greve (1998) noted, the changes may involve risk. In particular, in human service organizations, organizational members may fear that their jobs would be threatened if they do not meet performance goals or because they would not be able to adapt to newly adopted practices to improve performance (Proehl, 2001). Accordingly, when organizations are more risk averse, they are less willing to make the organizational changes even in the face of professionalization pressures. By following this line of logic, we expect:

H_{5b}: Risk aversion has a negative interaction on the relationship between professionalism and the adoption of outcome measurement.

Organizational Capacity

Following Eisinger (2002), we define organizational capacity as “a set of attributes that help or enable an organization to fulfill its missions” (p. 117). There has been significant scholarship aimed at understanding the effects of organizational capacity on performance measurement and management (e.g., Barman & MacIndoe, 2012; Berman & Wang, 2000; Franklin, 2002; Jordan & Hackbart, 1999; MacIndoe & Barman, 2012; Melkers & Willoughby, 2005). Jordan and Hackbart (1999) find that organizational capacity significantly influences the use of performance measures for budgeting and funding. Consistent with this, Berman and Wang (2000) suggest that the success of performance measurement depends on organizational capacity for undertaking performance measurement. In particular, the authors find that technical capacity and political capacity are significantly related to the breadth and depth of performance measurement. Their research is corroborated by Barman and MacIndoe (2012) arguing that organizational capacity is a vital component of outcome measurement. Their findings indicate that administrative capacity and technical capacity are significant predictors of the adoption of outcome measurement. To summarize, a wide array of literature highlights that organizational capacity is vital to successful performance measurement.

While many prior studies have addressed how organizational capacity is related to overall system of performance measurement, we focus on outcome measurement as a specific type of performance measurement. Further, given that outcome measurement has become increasingly important in human service nonprofits, as elaborated in the following section, it is important to explore the relationship between organizations' capacities and the adoption of outcome measurement, specifically in human service nonprofits.

Following past research, we classify organizational capacity into six dimensions: financial capacity, human resource capacity, technical capacity, administrative capacity, political capacity,

and network capacity (Barman & MacIndoe 2012; Berman & Wang, 2010; Christensen & Gazley, 2008; Hatry et al., 2003; MacIndoe & Barman, 2012; Poole, David, Reisman, & Nelson, 2001; Tuckman & Chang, 1991).

Financial capacity is an important capacity for undertaking outcome measurements. Hoefler (2000) surveyed Dallas nonprofits on why they did not conduct evaluations and revealed that about half of the respondents (48%) do not evaluate their programs because they do not have enough money to afford such evaluation. Nonprofits can employ performance measurement professionals, train their employees in outcome measurement, or develop a performance management system by which outcome information is appropriately used when they have sufficient financial resources for outcome measurement (MacIndoe & Barman, 2012). Therefore, we expect:

H_{6a}: Financial capacity is positively associated with the adoption of outcome measurement.

A second capacity nonprofits need to adopt for outcome measurements is staff that understands the outcome measurement process. To adopt outcome measurement, nonprofits need personnel who can collect outcome data, analyze the data, explain the outcome information, and relate it to their internal processes (Hatry et al., 2003). In a national survey of U.S. counties, however, only 64% of respondents reported that they have staff who can analyze performance data (Berman & Wang, 2000). As a primary resource provider, the government has greatly influenced outcome measurement practices of nonprofits (Barman & MacIndoe, 2012; Carman, 2009; Thompson & Riccucci, 1998). Hence, Berman and Wang (2000)'s finding has implications for performance measurement in the nonprofit sector. Complementary to this, Hoefler (2000) finds that almost half of the nonprofits (48%) did not evaluate their programs because "there was not enough staff time available to conduct an evaluation" and argues that the evaluation does not achieve its potential due to the insufficient organizational capacity (p. 171). Drawing upon the past research, we expect:

H_{6b}: Human resource capacity is positively associated with the adoption of outcome measurement.

A third capacity, technical capacity, refers to organizational infrastructure, information technology, and communication systems for meeting an organization's mission. In a national survey of U.S. counties, Berman and Wang (2000) find that 57.5% of respondents have relevant information technology for performance measurement, but only 29.1% can conduct surveys for measuring performance in a scientific way. The authors argue that technical capacity is significantly related to the widespread use of performance measurement. In the case of outcome measurement, nonprofits need adequate information technology and communication systems to collect, retain, analyze, and communicate their outcome measures. Therefore, we expect:

H_{6c}: Technical capacity is positively associated with the adoption of outcome measurement.

A fourth dimension of organizational capacity is administrative capacity. According to Barman and MacIndoe (2012), it is defined as "an organization's ability to implement an institutionally sanctioned practice such as outcome measurement" (p. 88). Adopting an organizational practice may involve a process to establish relevant procedures and administer them in a coordinated fashion. When organizations have administrative capacity, they can implement their practices in a consistent and coordinated way (Barman & MacIndoe, 2012; Weber, 1978). Thus, it is

Table 1. Survey Responses

NTEE	Email Survey (Qualtrics)			Pop-Up Contact Window		Total Number of Responses
	Number of Emails Sent	Number of Emails Bounced	Number of Responses	Number of Pop-Up Windows	Number of Responses	
P32	167	1	45	15	1	46
P43	434	2	148	33	8	156
P46	170	1	61	9	0	61
Sum	771	4	254	57	9	263

reasonable to expect that:

H_{6d}: Administrative capacity is positively associated with the adoption of outcome measurement.

A fifth capacity the literature has discussed is political capacity (e.g., Anguelovski & Carmin, 2011; Berman & Wang, 2010; Héritier, 2003). Berman and Wang (2010) highlight that counties with higher political capacity are more likely to use performance measurement. According to the authors’ research findings, counties with high political capacity are more likely to work with lower-level managers on their willingness to use performance measurement. Those counties are also more likely to undertake strategies to gain the support from elected officials on the need for performance measurement. When nonprofits adopt outcome measurement, political capacity helps establish a consensus to adopt the practice, convince stakeholders, assign funding for the new practice, and legitimize it (Berman & Wang, 2010; Héritier, 2003). Following this line of logic, we expect:

H_{6e}: Political capacity is positively associated with the adoption of outcome measurement.

Last, networking with other organizations and managing networks have been discussed in the literature as another dimension of organizational capacity (e.g., Christensen & Gazley, 2008; Eisinger, 2002; Moore, 2003). Obtaining all resources necessary for outcome measurement is difficult for a single nonprofit. By utilizing external networks, nonprofits may obtain various data, financial resources, competent employees, technology, and other assistance for adopting outcome measurement. Therefore, we expect:

H_{6f}: Network capacity is positively associated with the adoption of outcome measurement.

Methods and Data

This study tests the above hypotheses through survey data. The sample is drawn from the National Center for Charitable Statistics (NCCS) 2012 Core Data file because it contains data of all 501(c)(3) nonprofits required to file Form 990. In order to maintain the comparability of the organizations in the sample group, we limited our study to human service nonprofits. We further restrict the analysis to three sub-industries in the human service nonprofits—foster care organizations (NTEE code P32), family violence shelters (NTEE code P43), and family counseling organizations (NTEE code P46)—because the specific performance measures we ask about depend on particular industries. We also filtered out small nonprofit organizations with total expenses below \$500,000 to increase the likelihood of finding respondent organizations

Table 2. One Sample T-Test of Total Expenses

	<i>t</i> -test	Degree of Freedom	Sig. (2-tailed)	Mean Difference	95% Confidence Interval	
					Lower	Upper
Total Expenses	0.14	258	0.89	40,226.16	-533,828.57	614,280.90
Test Value = 2,642,202						

with a performance measurement system. With this sample screening, we obtained a total 1,039 human service nonprofits: 246 foster care organizations (P32), 569 family violence shelters (P43), and 224 family counseling organizations (P46). We collected contact information for each organization’s executive director or president from organizational websites. Through this process, we collected 775 email addresses of executive directors and sent the questionnaire through Qualtrics. Further, for organizations without a listed email for the executive director, we were able to use a pop-up contact window on 57 websites to send the survey to additional organizations. The survey was in the field from March 5 through April 26, 2014; 317 organizations consented to participate in our survey, providing a total 263 usable cases for data analysis. This gives us a final response rate of 31.76% (see table 1 for the details of the survey response rate).

To check the representativeness of the respondents’ group to the population, we performed one sample *t*-test with total expenses and chi-square test with regional distribution. The insignificant result of one sample *t*-test indicates that the average total expenses of respondent organizations are not significantly different from ones of population at a 95% confidence level (see table 2). In addition, cross tabulation and chi-square test of regional distribution indicate an insignificant result ($p=0.62$), which means there is no statistically significant difference in the regional distribution between population and respondents group (see table 3). Therefore, it can be concluded that the respondents group is representative of the population.

Measures: Dependent Variable

The Adoption of Outcome Measurement. To define and measure “adoption of outcome measurement,” many researchers take a dichotomous approach. For instance, Barman and MacIndoe (2012) asked survey respondents if more than half of their programs and services are subject to outcome measurement. Carman (2007) asked interviewees and survey respondents to describe the types of data (e.g., program expenditures, activities, outputs, and outcomes) they collect. Eckerd and Moulton (2010) also asked their survey respondents to indicate the types of evaluation method (e.g., program outputs, long-term program impact or outcome) adopted by their organizations. On the other hand, instead of taking this dichotomous approach, Thomson (2010) developed an index for measuring the extent of outcome measurement based on the number of expected outcomes, level of outcomes (short-term, intermediate, and end), and number of outcomes supported by organization’s evaluation tools.

We use a more robust approach than that of prior studies to measure the adoption of outcome measurement. We provided representative examples of outcome measures tailored to the nonprofit’s NTEE code and asked respondents how often they collect the information on each outcome. To obtain the field-specific outcomes, we searched the American Association for Marriage and Family Therapy (AAMFT, 2013), Child and Family Services Agency (CFSA, 2012), and nonprofit organizations’ websites. As a professional association of more than 25,000 marriage and family therapists in the United States, Canada, and abroad, AAMFT has established the professional standards of the marriage and family therapy. CFSA, a public child welfare agency, has developed performance indicators to measure the well-being of children and

Table 3. Crosstab and χ^2 Test of Regional Distribution

Region	Respondents	Population	Total
Midwest	72 (27.8%)	260 (25.0%)	332 (25.6%)
Northeast	40 (15.4%)	189 (18.2%)	229 (17.6%)
Puerto Rico	0 (0.0%)	4 (0.4%)	4 (0.3%)
South	82 (31.7%)	317 (30.5%)	399 (30.7%)
West	65 (25.1%)	269 (25.9%)	334 (25.7%)
Total	259 (100.0%)	1039 (100.0%)	1298 (100.0%)

In $n(\%)$, n indicates number and % indicates proportion.
 Pearson $\chi^2(4) = 2.648, p = 0.618$

families. As to the nonprofits' websites, we selected well-developed websites showing the programs, services, and performance indicators of the organization. In addition to the example outcomes, we added an open-ended question so that respondents could enter the outcomes they are measuring but not listed in our survey. Last, drawing on MacIndoe and Barman (2012, p. 9), we asked one more question: "In the last fiscal year, what percentage of your organization's programs/services were subject to outcome measurement?" on a five-point scale. The responses to these two questions related to our dependent variable were compared in order to check the reliability of the responses.

To filter unreliable responses, we compared the responses of the question regarding the percentage of programs/services subject to outcome measurement and the count of outcome measures: 120 out of the 263 respondents answered the highest value in the percentage of outcome measurement. However, one out of the 120 respondents neither checked any outcome measure listed nor provided any additional outcome measures. In addition, we checked the cases where respondents did not check any outcome measure nor provide any additional outcome measures. We assumed that these respondents would answer the lowest value in the percentage question and found that two more responses are not reliable. Last, we found that one respondent did not answer the percentage question while he/she answered the question of example outcomes. In brief, four responses did not pass our data quality examination, and we excluded these cases from the data analysis.

As previously described, we presented a list of outcome measures tailored to each NTEE code and counted how many outcomes each respondent measures. Different from our expectation, we found very small variation in the responses and negative skewness in the distribution, which means that a large percentage of respondents checked more than 50% of the outcome measures listed in the survey. Given the skewness in our data, we collapsed the categories with low frequencies and recoded the dependent variable into dichotomous variable (see table 4).

Measures: Independent Variables

In order to avoid measuring the exact same phenomena with our dependent and independent variables, we construct our independent variables to capture overall organization attitudes toward risk, institutional pressures, and capacities. For example, if we only asked about organizational capacities specifically related to outcome measurement, we would capture the

Table 4. Dependent Variable

	P32	P43	P46	Total
Less than 50%	22% (10)	24% (37)	33% (20)	26% (67)
More than 50%	78% (35)	76% (117)	67% (40)	74% (192)
Total	100% (45)	100% (154)	100% (60)	100% (259)

n is reported in parentheses

extent to which they have allocated/tailored organizational capacities to an outcome measurement system. Essentially, this would measure the extent to which they have adopted outcome measurement, which is our dependent variable. Because an outcome measurement system is but one of many management systems in a nonprofit, our strategy for measuring our independent variables provides us with insights into how overall organizational characteristics predict the likelihood of a nonprofit adopting an outcome measurement system.

Risk Aversion. This variable is measured by asking three survey items: 1.) “There is a reward/incentive system in place in the organization that encourages risk taking (using innovative ideas with the goal of improving performance)”; 2.) “Management is willing to implement appropriate organizational innovation and change”; and, 3.) “Nonmanagement employees willingly accept organizational innovation and change” (Julnes & Holzer, 2001, p. 706). These survey items are scored on a five-point Likert scale, and our overall scale of risk aversion is generated by averaging the response of the three items. The Cronbach alpha for the three items scale is 0.58.

In addition, we measured some of our independent variables (e.g., mimetic isomorphism, political capacity, network capacity) with five-point Likert scales. Because we found very small variation in the responses, we collapsed the categories with small frequencies and recoded the five-point categorical variables into dichotomous variables. Values of 4 and 5 were recoded into 1 and those of 1, 2, and 3 were recoded into 0.

Institutional Factors. First, *coercive isomorphism* is measured by asking if the nonprofit organization receives funding from the United Way (Barman & MacIndoe, 2012). As one of the first national agencies to request its network members to make a distinction between outcomes and outputs, the United Way has promoted the adoption of outcome measurement in the United States since the early 1990s (Ebrahim & Rangan, 2010) and requires its grantees to conduct outcome measurement (Hendricks, Plantz, & Pritchard, 2008). This can exert coercive pressure on the grantee organizations. Second, *mimetic isomorphism* is measured by asking “The largest organizations in our field measure their outcomes” on a five-point scale (Barman & MacIndoe, 2012). Next, we classify *uncertainty* into two dimensions: socioeconomic uncertainty and political uncertainty (Andrews, 2008). Survey items for measuring socioeconomic uncertainty and political uncertainty are “Changes in our community’s socioeconomic status are predictable” (reversed) (Andrews, 2008; Guo & Acar, 2005; Hoque, 2004) and “Changes in our community’s political environment are predictable” (reversed) (Andrews, 2008; Budding, 2004; Hoque, 2004) respectively. These two items are measured on a five-point scale and averaged to create the overall scale of uncertainty. Cronbach’s alpha of the two items is 0.77, indicating an adequate level of reliability. Last, to measure *professionalism*, as with past scholarship (e.g., Hildebrandt & Eom, 2011; Mikkelsen, 2013; Ullian & Schink, 2012), we ask if the nonprofit is accredited. The survey item for the professionalism asks, “Is your organization accredited by any external bodies?”

Table 5. Descriptive Statistics

Variables	N	Mean	Min	Max	SD
Adoption of Outcome Measurement	259	0.74	0	1	
Risk Aversion	250	2.11	1	5	0.69
United Way Funding	245	0.63	0	1	
Largest Organization's Outcome Measurement	233	0.79	0	1	
Uncertainty	229	3.00	1	5	0.88
Accreditation	240	0.46	0	1	
Financial Capacity (ER)	259	0.73	-3.25	1.00	0.42
Human Capacity	253	0.88	0	1	
Technical Capacity	257	0.73	0	1	
Administrative Capacity	258	0.98	0	1	
Political Capacity	246	0.69	0	1	
Network Capacity	252	0.77	0	1	
Organizational Age	258	37.10	8	201	23.91
Total Expenses	259	2,682,428	508,178	46,152,548	4,691,514

Organizational Capacity Factors. As described in the hypotheses section, our analysis includes six elements of organizational capacity. First, we measure *financial capacity*, using Bowman (2011)'s long-term financial capacity, equity ratio (ER), which is calculated by the following equation:

$$ER = \frac{\text{Total Assets} - \text{Total Liabilities}}{\text{Total Assets}}$$

Second, *human resource capacity* is measured, using a survey item, "Our organization has staff with the skills and expertise necessary to track program participants" (Poole et al., 2001). Third, the survey item for measuring *technical capacity* is "Our organization has adequate infrastructure (e.g., information technology, communication system) for daily operations" (Berman & Wang, 2010). Fourth, following prior research that administrative capacity derives from bureaucratic dimensions, including written rules regarding organizational administration (Hall, 1968; Weber, 1978), we measure *administrative capacity* by asking "Our organization has written rules concerning compensation, governance, job descriptions, conflict of interest, and/or volunteers" (Barman & MacIndoe, 2012). Fifth, as for *political capacity*, while it can comprise diverse elements, we focus on bargaining and negotiation abilities because adopting outcome measurement may involve a process for engaging and negotiating with stakeholders. We measure political capacity through the survey item, "Our organization frequently discusses operational issues with stakeholders" (Héritier, 2003). Lastly, given that outcome measurement may require various resources from external help networks, *network capacity* is measured by asking "Our organization seeks help from other organizations to improve decision making" (Christensen & Gazley, 2008; Eisinger, 2002). Except financial capacity, which is a continuous variable, the dimensions of organizational capacity are measured on five point scales, with higher scores indicating higher capacities.

Control Variables. *Organizational size* and *age* are considered here as control variables. *Organizational size* is measured, using total expenses, and *organizational age* is measured by asking survey item "In which year was your organization founded?"

Descriptive statistics of the variables are reported in table 5.

Table 6. Logistic Regression Model

Independent Variables	β	S.E.	Sig.	Odds Ratio
NTEE(1) – P32	.986	.592	.096	2.682
NTEE(2) – P43	.724	.437	.097	2.064
Risk Aversion	.970	.555	.081	2.639
United Way Funding	1.179**	.406	.004	3.251
Largest Organization’s Outcome Measurement	4.621*	2.107	.029	101.618
Largest Organization’s Outcome Measurement*Risk Aversion	-.993	.581	.087	.371
Uncertainty	.517	.411	.209	1.678
Uncertainty*Largest Organization’s Outcome Measurement	-.765	.493	.123	.465
Accreditation	3.361**	1.175	.004	28.821
Accreditation*Risk Aversion	-1.348**	.512	.008	.260
Financial Capacity	-.367	.443	.407	.693
Human Resource Capacity	1.046*	.518	.044	2.847
Infrastructure Capacity	-.699	.441	.113	.497
Administrative Capacity	-.704	1.159	.543	.494
Political Capacity	.765*	.380	.044	2.148
Network Capacity	-.173	.411	.674	.841
Ln(Organizational Age)	.432	.382	.257	1.541
Ln(Total Expenses in 2011)	.448*	.221	.043	1.565
Constant	-11.711	3.847	.002	.000

Reference category: P46

This logistic model was obtained from the imputed data.

**p < .01; *p < .05 (two-tailed tests)

Missing Data

Since data analysis can be distorted by improper handling of missing data, we performed a missing value analysis. We found that overall, the percentage of incomplete values in our dataset is 4.54%. However, 41.31% of the cases (107 out of total 259 responses) contains missing values. Missing data is non-ignorable when the number of cases with missing information is larger than 5% (Garson, 2012a). Hence, following past research, we undertook multiple imputation to deal with the missing data problem (Graham, Hofer, Donaldson, MacKinnon, Schafer, 1997; Graham & Schafer, 1999; Schafer & Graham, 2002; Wayman, 2003). Multiple imputation is not only simpler than other methods for dealing with missing data problem but also can produce unbiased parameter estimates, in particular, when the rate of missing data is high (Wayman, 2003). In our study, multiple imputation procedure generated five imputed datasets and we used regression model derived from pooled dataset.

Results

Since we recoded the dependent variable into a binary variable, we used logistic regression for analyzing our data. In order to find the best model for the data analysis, we examined our models based on five attributes of a good model: parsimony, identifiability, goodness of fit, theoretical consistency, and predictive power (Gujarati & Porter, 2010). The best model is provided in table 6.

As reported in table 6, coercive isomorphism (united way funding), mimetic isomorphism (largest organizations' outcome measurement), professionalism (accreditation), human resource capacity, and political capacity have significant results. In addition, the interaction term between risk aversion and professionalism is significant. Last, total expenses as a control variable are statistically significant. These results are discussed in the subsequent section based on our hypotheses.

Hypothesis 1 (Risk Aversion)

Hypothesis 1 examines how risk aversion is associated with the adoption of outcome measurement. The logistic regression results show that the main effect of risk aversion is not significant. However, the interaction term between risk aversion and accreditation is negatively significant at a 0.05 level ($p=0.01$). These results mean that, although an organization with accreditation (see the evaluation of hypothesis 5a below) is more likely to adopt outcome measurement, the positive impact of accreditation on the adoption of outcome measurement is attenuated when the organization is more risk averse. These results provide partial support for Hypothesis 1.

Hypothesis 2 (Coercive Isomorphism)

Hypothesis 2 tests the effect of coercive isomorphism on the adoption of outcome measurement. The positive and statistically significant ($p=0.00$) coefficient on the "United Way funding" variable provides support for this hypothesis. As shown in table 6, the odds ratio of "United Way funding" indicates that nonprofits with United Way funding are 3.25 times more likely to adopt the outcome measurement, compared with the nonprofits without United Way funding.

Hypothesis 3a and 3b (Mimetic Isomorphism)

Hypothesis 3a predicts a positive influence of mimetic isomorphism on the adoption of outcome measurement. The regression results support this hypothesis. The "largest organizations use outcome measurement" variable is positively significant ($p=0.03$) in the model. The odds ratio in table 6 shows that nonprofits with more mimetic isomorphism are 101.62 times more likely to adopt outcome measurement, compared with the nonprofits with less mimetic isomorphism. On the other hand, Hypothesis 3b, which expects a positive interaction effect between mimetic isomorphism and risk aversion, is not supported by the findings. One explanation could be that adopting a similar behavior in pursuit of risk minimization is common when an organization's environment is highly competitive. To put it another way, mimetic behavior can alleviate the risk of failure by preserving "the status quo among competitors" when competitive rivalry exists or competition is intense (Lieberman & Asaba, 2006, p. 367). On the other hand, because nonprofits face less competitive environments compared with private firms, the positive interaction effect between risk aversion and mimetic behavior predicted by theory may be more applicable to for-profit organizations.

Hypothesis 4 (Uncertainty)

Hypothesis 4 examines how uncertainty influences the relationship between mimetic isomorphism and the adoption of outcome measurement. Unlike our expectation, however, the logistic regression results indicate that the influence of uncertainty is insignificant ($p=0.12$). This unexpected result can be explained by considering two different approaches to uncertainty. On the one hand, institutional theorists argue that mimetic behavior results from "standard responses to uncertainty" (DiMaggio & Powell, 1983, p.150). Drawing upon this institutional

Table 7. Adjusted B Coefficient and Odds Ratio of Accreditation

Risk Aversion	Adjusted b	Odds Ratio
1	2.013	7.486
2	0.665	1.944
3	-0.683	0.505
4	-2.031	0.131
5	-3.379	0.034

theory approach, we expected that uncertainty would positively influence the relationship between mimetic isomorphism and the adoption of outcome measurement. On the other hand, agency theorists maintain that uncertainty may create unexpected outcomes; thus, it is difficult for principals to correctly measure the outcomes of their agents in an uncertain environment (e.g., Celly & Frazier, 1996). Further, because the outcomes may fluctuate widely in an uncertain environment, it is difficult to control and measure the outcomes in that situation (Kumar, Scheer, & Steenkamp, 1995), which may negatively influence the adoption of outcome measurement. In this regard, agency theorists argue that uncertainty might be negatively associated with the adoption of outcome measurement. In brief, these two conflicting effects of uncertainty on the adoption of outcome measurement might cancel each other out, thus resulting in an insignificant effect.

Hypothesis 5a and 5b (Normative Isomorphism)

Hypothesis 5a tests the positive effect of professionalism on the adoption of outcome measurement, which is supported by the regression results, as “accreditation” is positively significant ($p=0.00$). Before interpreting the odds ratio of accreditation, we adjusted the b coefficient of accreditation because accreditation is involved in an interaction term with risk aversion (Hypothesis 5b) as follows.

$$\text{Adjusted } \beta \text{ Coefficient of Accreditation} = 3.36 - 1.35 * \text{Risk Aversion}.$$

We interpret the odds ratio of accreditation, following the approach of past research (e.g., Garson, 2012b; Li & Barry, 2012). As shown in table 7, the b coefficients of accreditation are positive when the risk aversion is at the low level, whereas they become negative at the high level of risk aversion. To put it another way, the positive influence of accreditation on the adoption of outcome measurement decreases as the risk aversion level increases. When risk aversion is at the lowest level, nonprofits with accreditation are 7.49 times more likely to adopt the outcome measurement than otherwise. On the other hand, when risk aversion increases 1 unit, nonprofits with accreditation are only 1.94 times more likely to adopt outcome measurement than otherwise.

Hypothesis 6a (Financial Capacity)

Unlike our expectation, Hypothesis 6a is not supported by the regression results, as financial capacity is not significant ($p=0.41$). A potential explanation for this finding is that small nonprofits with total expenses below \$500,000 were excluded from our analysis; thus, we did not capture the full range of variation in the financial capacity of the population group. Additionally, there might be a bias in our sample. Table 8 shows that the financial capacity (ER) of the respondents group is higher than the entire population’s, even though total expenses do

Table 8. ER of Respondent and Population Groups

NTEE	Respondent group (n = 259)			Population group (n = 1039)		
	Mean	SD	Median	Mean	SD	Median
P32	0.51	0.73	0.75	0.31	2.34	0.67
P43	0.81	0.20	0.88	0.77	0.49	0.88
P46	0.68	0.46	0.84	0.62	0.48	0.77
Total	0.73	0.42	0.84	0.63	1.24	0.82

not differ. Through an independent t-test, we confirmed that this difference is statistically significant with 95% confidence level. Nonprofits with higher capacity to answer surveys may also be more likely to have the organizational resources for outcome measurement.

Hypothesis 6b (Human Resource Capacity)

Hypothesis 6b is supported by the regression result as human resource capacity is positively significant ($p=0.04$). As shown in the odds ratio (see table 6), nonprofits with higher human resource capacity are 2.85 times more likely to adopt outcome measurement, compared with the nonprofits with lower human resource capacity.

Hypothesis 6c (Technical Capacity)

Hypothesis 6c tests the influence of technical capacity on the adoption of outcome measurement. Unlike our expectations, however, Hypothesis 6c is not supported by these data; technical capacity is not significant ($p=0.11$). Prior studies provide a potential explanation for this finding. In our study, technical capacity is measured by asking if the organization has adequate infrastructure (e.g., information technology) for daily operations, and a majority of the respondents (73%) agreed with this survey item in our study. However, prior studies (e.g., Corder, 2001; Saigal, 2008) argue that nonprofits do not have sufficient training to make use of their infrastructure. Even if nonprofits have adequate infrastructure for outcome measurement, it may not directly influence the adoption of outcome measurement if the staff within the agency has not received a training for the effective use of that infrastructure.

Hypothesis 6d (Administrative Capacity)

The influence of administrative capacity on the adoption of outcome measurement is not supported by the regression result ($p=0.54$). Very little variation in administrative capacity may explain this non-finding. Because more than 98% of respondents agreed with the survey item measuring administrative capacity, there is not much variation in this variable, meaning the effect size would need to be quite large to be statistically significant. Insufficient variation in administrative capacity might be the reason that its influence is insignificant in the analysis results.

Hypothesis 6e (Political Capacity)

Hypothesis 6e examines the effect of political capacity on the adoption of outcome measurement. Because the political capacity is positively significant in the logistic regression model at the 0.05 significance level ($p=0.04$), we conclude that Hypothesis 6e is supported by our findings. Table 6 indicates that nonprofits with higher political capacity are 2.15 times more likely to adopt outcome measurement, compared with the nonprofits with lower political capacity.

Hypothesis 6f (Network Capacity)

Hypothesis 6f, which examines the effect of network capacity on the adoption of outcome measurement, is not supported by the regression result ($p=0.67$). The characteristics of human service nonprofit organizations may explain this unexpected finding. Public organizations have a comparatively homogenous network of resources, whereas human service nonprofits have a more diverse network of resources (Corder, 2001). Cohen and Levinthal (1990) argue that human service nonprofits need capacity to manage this diverse network of resources. However, we did not capture this aspect in the survey item for measuring the network capacity. Consideration of the unique characteristics of human service nonprofits, including the complexity of networks, may result in different findings.

Discussion

Our findings indicate that each of the three theory bases we examine in this study—agency theory, institutional theory, and organizational capacity—contribute to our understanding of what facilitates (or inhibits) nonprofits adopting outcome measurements. We explore these findings in more depth below.

Agency Theory – Risk Aversion

Although there are a few studies regarding performance measurement, accountability, and issues of risk in nonprofit practice, (e.g., Benjamin, 2008), there remains limited attention paid to the effect of an organization's risk attitude on the adoption of performance measurement, focusing on outcome rather than other types of measures. Our study fills this gap by exploring the main effect and interactions of risk aversion on the adoption of outcome measurements. Agency theorists assume that principals can diversify their investments, whereas agents cannot diversify their employment; thus, they are risk averse (e.g., Davis et al., 1997; Eisenhardt, 1989). This classic assumption has been extended by subsequent organization theorists and researchers to allow risk attitudes within this theory to vary widely (e.g., Eisenhardt, 1989; MacCrimmon & Wehrung, 1986; Shapiro, 2005). Our study provides empirical evidence in favor of this more expansive approach but not in a straightforward way. Based on our findings, we do not discover a direct relationship between risk aversion and the adoption of outcome measurement. However, we do see that some of the variation in the adoption of outcome measurement can be explained by nonprofits' risk aversion interaction with other factors (e.g., institutional pressures). This finding has implications for researchers concerned with organization risk. Given that risk redistribution between principals and agents is the main issue that agency theory is concerned with, the broader organizational context must be considered to understand the impact of risk aversion on organizational decisions such as whether to adopt outcome measures. Future research should delve into the other applications of agency theory assumptions (e.g., the length of agency relationship, task programmability, and information system) as well as risk aversion not only as it relates to outcome measurement, but other organization behaviors as well.

Institutional Factors

From an institutional theory perspective, performance measurement is a largely symbolic (or ceremonial) practice to acquire institutional legitimacy by conforming to institutional norms (e.g., Eckerd & Moulton, 2010; Lee & Nowell, 2015; Modell, 2009; Roy & Seguin, 2000; Yang, 2009). In our study, we found that coercive, mimetic, and normative pressures have positive

and significant influences on the adoption of outcome measurement. According to Roy and Seguin (2000), the adoption of organizational practices for improving performance can make the organization appear proactive, innovative, and effective, even if it is only ritual and ceremonial.

As shown in table 6, United Way funding is significantly related to the adoption of outcome measurement. Because the United Way is not the only grantor that imposes outcome measurement (a potential construct validity issue that is discussed with other limitations below), measuring coercive isomorphism from this one source might underestimate the true effect of coercive isomorphism on the adoption of outcomes measurement. However, the significant influence of United Way funding on outcome measurement corroborates the notion that the coercive pressure exerted on grantee organizations can shape their organizational practices.

In addition, we identified a significant relationship between mimetic isomorphism and the adoption of outcome measurement. In Barman and MacIndoe (2012)'s findings, mimetic isomorphism was insignificantly related to outcome measurement. The authors justified this result by arguing that mimetic isomorphism occurs only when the practice is widespread in the industry (Tolbert & Zucker, 1983). Whereas outcome measurement might not be common across the entire nonprofit sector, the sample used in Barman & MacIndoe's (2012) study, our focus on human service nonprofits, where outcome measurement is comparatively more widespread, may explain our significant result.

Finally, our study demonstrated the positive influence of normative isomorphism on the adoption of outcome measurement in nonprofits. The growing body of scholarship dedicated to understanding outcome measurement (e.g., Barman & MacIndoe, 2012; Eckerd & Moulton, 2010; Julnes & Holzer, 2001; MacIndoe & Barman, 2012; Melkers & Willoughby, 2005; Thomson, 2010) appears to be having an impact on the education and practice of nonprofit practitioners. Future research needs to explore whether this adoption of outcome measurements pushed by scholarship and pedagogy is only symbolic, or if it is actually used in meaningful ways to improve organizational practices.

Organizational Capacity

By identifying the variation in human resource capacity across nonprofits and its significant effect, our study confirmed the importance of staff trained to track program participants. Capturing outcomes requires similar skills and knowledge. Our study empirically demonstrates that staff who can track program participants and analyze that data are needed for adopting outcome measurement. In addition to human resource capacity, we found a significant influence of political capacity on the adoption of outcome measurement. Although Berman and Wang (2000) already examined the effect of political capacity on performance measurement, their research was focused on public organizations. By exploring the effect of political capacity in the case of nonprofit organizations and considering other perspectives, our study both corroborates the research conducted by Berman and Wang (2000) and extends its implications to nonprofit organizations.

Interestingly, the other four dimensions of organizational capacity did not show significant results. According to Julnes and Holzer (2001), however, there are two different stages in performance measurement—adoption and implementation—which may have different mechanisms. Future researchers exploring the actual uses (e.g., outcome information uses in the strategic planning process) of outcome measures may have different findings.

Conclusion

In this article, we used agency theory, institutional theory, and organizational capacity to explore the factors related to variation in the adoption of outcome measurement. Based on survey data from 259 human service nonprofits, we found that nonprofit organizations are more likely to adopt outcome measurement when they are less risk averse; influenced by coercive, mimetic, and normative pressures; and have higher human resource and political capacities.

Although our findings provide meaningful implications for practice and theory, we note some limitations of our study. First, our study limits the generalizability of the findings. As discussed in the methods section, we excluded small nonprofits with total expenses below \$500,000 in the sampling process and focused on mid- and large-sized nonprofits equipped with more skills and knowledge to develop performance measurement systems. While this filtering increases the probability finding nonprofits with a performance measurement system, it limits the generalizability of our findings to all sizes of nonprofits. Another limitation lies with the use of single item measures (e.g., coercive isomorphism—United Way funding, mimetic isomorphism—the largest organizations' outcome measurement, normative isomorphism—accreditation, financial capacity—ER), which could be criticized in terms of construct validity and scale reliability. In addition to the use of single-item measures, our study has a conceptualization issue. In operationalizing organizational attitude toward risk, we mainly focused on risk-taking inside an organization. Given that the certainty of the task environment also may affect an organization's decision to contract on outcomes and adopt outcome measurement, however, risk-taking in the task environment also needs to be considered. Finally, our 259 cases also limit the number of variables we can use in multivariate modeling. Future researchers may have different findings by using multiple measures of a single construct, collecting a sufficient number of cases (achieving a higher survey response rate), and resolving the conceptualization issue in our study.

Our research makes several contributions to the literature despite its limitations. First, although excellent work on outcome measurement has already been done (e.g., Eckerd & Moulton, 2010; Julnes & Holzer, 2001; MacIndoe & Barman, 2012; Thomson, 2010), our study discussed factors that have received comparatively limited attention in current scholarship (e.g., risk aversion, network capacity). Second, much of the prior research on the outcome measurement were grounded in analyses of particular regions (e.g., Eckerd & Moulton, 2010; MacIndoe & Barman, 2012; Thomson, 2010; Zimmermann & Stevens, 2006), which limits the generalizability of the findings. On the other hand, because our study is not limited to specific regions, it is more generalizable to human services nonprofits across the country compared with the prior research focused on one city or state. Last, since we focused on specific NTEE codes—P32, P43, and P46—we could ask about collecting specific outcome measures rather than just asking about outcome measurement generically in the survey. This strategy enhanced the reliability and validity of our survey results for U.S. nonprofits operating in those three human services fields.

Finally, while this current research focuses on the adoption of outcome measurement, the implementation of performance measurement (or actual use of performance data) needs to be examined in future research. While our findings related to institutional theory provide some evidence for the theoretical assertion that outcome measurement may be adopted for ceremonial purposes, additional study is needed to explore whether and how these systems are used in organizations. This is a fertile area for future research that has additional important implications for whether nonprofit organizations can advance their practices by better understanding and using their performance data to achieve their missions.

Disclosure Statement

The authors declare that there are no conflicts of interest that relate to the research, authorship, or publication of this article.

References

- Aldrich, H. E., & Fiol, C. M. (1994). Fools rush in? The institutional context of industry creation. *Academy of Management Review*, *19*, 645-670. [doi:10.2307/258740](https://doi.org/10.2307/258740)
- Amaratunga, D., & Baldry, D. (2002). Moving from performance measurement to performance management. *Facilities*, *20*, 217-223. [doi:10.1108/02632770210426701](https://doi.org/10.1108/02632770210426701)
- American Association for Marriage and Family Therapy. (2013). *Marriage and family therapists: The friendly mental health professionals*. Retrieved from [http://www.aamft.org/imis15/content/consumer_updates/Marriage and Family Therapists.aspx](http://www.aamft.org/imis15/content/consumer_updates/Marriage_and_Family_Therapists.aspx)
- Andrews, R. (2008). Perceived environmental uncertainty in public organizations: An empirical exploration. *Public Performance and Management Review*, *32*, 25-50. [doi:10.2753/PMR1530-9576320102](https://doi.org/10.2753/PMR1530-9576320102)
- Angelovski, I., & Carmin, J. (2011). Something borrowed, everything new: Innovation and institutionalization in urban climate governance. *Current Opinion in Environmental Sustainability*, *3*, 169-175. [doi:10.1016/j.cosust.2010.12.017](https://doi.org/10.1016/j.cosust.2010.12.017)
- Barman, E., & MacIndoe, H. (2012). Institutional pressures and organizational capacity: The case of outcome measurement. *Sociological Forum*, *27*, 70-93. [doi:10.1111/j.1573-7861.2011.01302.x](https://doi.org/10.1111/j.1573-7861.2011.01302.x)
- Benjamin, L. M. (2008). Bearing more risk for results performance accountability and nonprofit relational work. *Administration and Society*, *39*, 959-983. [doi:10.1177/0095399707309357](https://doi.org/10.1177/0095399707309357)
- Berman, E., & Wang, X. (2000). Performance measurement in US counties: Capacity for reform. *Public Administration Review*, *60*, 409-420. [doi:10.1111/0033-3352.00104](https://doi.org/10.1111/0033-3352.00104)
- Bohte, J., & Meier, K. J. (2000). Goal displacement: Assessing the motivation for organizational cheating. *Public Administration Review*, *60*, 173-182. [doi:10.1111/0033-3352.00075](https://doi.org/10.1111/0033-3352.00075)
- Bolman, L. G., & Deal, T. E. (1991). *Reframing organizations: Artistry, choice and leadership*. San Francisco, CA: Jossey-Bass.
- Bowman, W. (2011). Financial capacity and sustainability of ordinary nonprofits. *Nonprofit Management and Leadership*, *22*, 37-51. [doi:10.1002/nml.20039](https://doi.org/10.1002/nml.20039)
- Budding, G. T. (2004). Accountability, environmental uncertainty and government performance: Evidence from Dutch municipalities. *Management Accounting Research*, *15*, 285-304. [doi:10.1016/j.mar.2004.04.003](https://doi.org/10.1016/j.mar.2004.04.003)
- Cairns, B., Harris, M., Hutchison, R., & Tricker, M. (2005). Improving performance? The adoption and implementation of quality systems in UK nonprofits. *Nonprofit Management and Leadership*, *16*, 135-151. [doi:10.1002/nml.97](https://doi.org/10.1002/nml.97)
- Carman, J. G. (2007). Evaluation practice among community-based organizations research into the reality. *American Journal of Evaluation*, *28*, 60-75. [doi:10.1177/1098214006296245](https://doi.org/10.1177/1098214006296245)
- Carman, J. G. (2009). Nonprofits, funders, and evaluation accountability in action. *American Review of Public Administration*, *39*, 374-390. [doi:10.1177/0275074008320190](https://doi.org/10.1177/0275074008320190)
- Carman, J. G., & Fredericks, K. A. (2008). Nonprofits and evaluation: Empirical evidence from the field. *New Directions for Evaluation*, *2008*, 51-71. [doi:10.1002/ev.268](https://doi.org/10.1002/ev.268)
- Carrilio, T. E., Packard, T., & Clapp, J. D. (2004). Nothing in—nothing out: Barriers to the use of performance data in social service programs. *Administration in Social Work*, *27*, 61-75. [doi:10.1300/J147v27n04_05](https://doi.org/10.1300/J147v27n04_05)

- Celly, K. S., & Frazier, G. L. (1996). Outcome-based and behavior-based coordination efforts in channel relationships. *Journal of Marketing Research*, 33, 200-210. doi:10.2307/3152147
- Children and Family Services Agency. (2012). *FY 2012 performance plan*. Retrieved from <http://oca.dc.gov/sites/default/files/dc/sites/oca/publication/attachments/CFSA12.pdf>
- Christensen, R. K., & Gazley, B. (2008). Capacity for public administration: Analysis of meaning and measurement. *Public Administration and Development*, 28, 265-279. doi:10.1002/pad.500
- Cohen, W., & Levinthal, D. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128-152. doi:10.2307/2393553
- Corder, K. (2001). Acquiring new technology comparing nonprofit and public sector agencies. *Administration and Society*, 33, 194-219. doi:10.1177/00953990122019730
- Dart, R. (2004). The legitimacy of social enterprise. *Nonprofit Management and Leadership*, 14, 411-424. doi:10.1002/nml.43
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). Toward a stewardship theory of management. *Academy of Management Review*, 22, 20-47. doi:10.2307/259223
- Dicke, L. A. (2002). Ensuring accountability in human services contracting can stewardship theory fill the bill? *American Review of Public Administration*, 32, 455-470. doi:10.1177/027507402237870
- Dicke, L. A., & Ott, J. S. (1999). Public agency accountability in human services contracting. *Public Productivity and Management Review*, 22, 502-516. doi:10.2307/3380933
- DiMaggio, P., & Powell, W. W. (1983). The iron cage revisited: Collective rationality and institutional isomorphism in organizational fields. *American Sociological Review*, 48, 147-160. doi:10.2307/2095101
- Ebrahim, A. S., & Rangan, V. K. (2010). *The limits of nonprofit impact: A contingency framework for measuring social performance* (Working Paper No. 10-099). Cambridge, MA: Harvard Business School General Management Unit.
- Eckerd, A., & Moulton, S. (2010). Heterogeneous roles and heterogeneous practices: Understanding the adoption and uses of nonprofit performance evaluations. *American Journal of Evaluation*, 32, 98-117. doi:10.1177/1098214010381780
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14, 57-74. doi:10.5465/AMR.1989.4279003
- Eisinger, P. (2002). Organizational capacity and organizational effectiveness among street-level food assistance programs. *Nonprofit and Voluntary Sector Quarterly*, 31, 115-130. doi:10.1177/0899764002311005
- Ellis, S., & Shpielberg, N. (2003). Organizational learning mechanisms and managers' perceived uncertainty. *Human Relations*, 56, 1233-1254. doi:10.1177/00187267035610004
- Fligstein, N. (1985). The spread of the multidivisional form among large firms, 1919-1979. *American Sociological Review*, 50(3), 377-391.
- Franklin, A. L. (2002). An examination of the impact of budget reform on Arizona and Oklahoma appropriations. *Public Budgeting and Finance*, 22, 26-45. doi:10.1111/1540-5850.00079
- Frumkin, P., & Galaskiewicz, J. (2004). Institutional isomorphism and public sector organizations. *Journal of Public Administration Research and Theory*, 14, 283-307. doi:10.1093/jopart/muh028
- Garson, G. David. (2012a). *Data imputation for missing values*. Asheboro, NC: Statistical Associates Publishing.
- Garson, G. David. (2012b). *Logistic regression*. Asheboro, NC: Statistical Associates Publishing.
- Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and*

- substance abuse research* (pp. 325-366). Washington, DC: American Psychological Association. [doi:10.1037/10222-010](https://doi.org/10.1037/10222-010)
- Graham, J. W., & Schafer, J. L. (1999). On the performance of multiple imputation for multivariate data with small sample size. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 1-27). Thousand Oaks, CA: Sage.
- Greve, H. R. (1998). Performance, aspirations, and risky organizational change. *Administrative Science Quarterly*, *43*, 58-86. [doi:10.2307/2393591](https://doi.org/10.2307/2393591)
- Gujarati, D., & Porter, D. (2010). *Essentials of econometrics*. New York, NY: McGraw-Hill/Irwin.
- Guo, C., & Acar, M. (2005). Understanding collaboration among nonprofit organizations: Combining resource dependency, institutional, and network perspectives. *Nonprofit and Voluntary Sector Quarterly*, *34*, 340-361. [doi:10.1177/0899764005275411](https://doi.org/10.1177/0899764005275411)
- Hall, R. H. (1968). Professionalization and bureaucratization. *American Sociological Review*, *33*, 92-104. [doi:10.2307/2092242](https://doi.org/10.2307/2092242)
- Hatry, H., Lampkin, L., Morley, E., & Cowan, J. (2003). *How and why nonprofits use outcome information*. Washington, DC: The Urban Institute.
- Haveman, H. A. (1993). Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly*, *38*, 593-627. [doi:10.2307/2393338](https://doi.org/10.2307/2393338)
- Hendricks, M., Plantz, M. C., & Pritchard, K. J. (2008). Measuring outcomes of United Way-funded programs: Expectations and reality. *New Directions for Evaluation*, *2008*, 13-35. [doi:10.1002/ev.266](https://doi.org/10.1002/ev.266)
- Héritier, A. (2003). New modes of governance in Europe: Increasing political capacity and policy effectiveness. In T. A. Borzel & R. A. Cichowski (Eds.), *The state of the European Union: Vol. 6. Law, politics, and society* (pp. 105-126). Oxford, England: Oxford University Press.
- Hildebrandt, S. A., & Eom, M. (2011). Teacher professionalization: Motivational factors and the influence of age. *Teaching and Teacher Education*, *27*, 416-423. [doi:10.1016/j.tate.2010.09.011](https://doi.org/10.1016/j.tate.2010.09.011)
- Hoefler, R. (2000). Accountability in action? Program evaluation in nonprofit human service agencies. *Nonprofit Management and Leadership*, *11*, 167-177. [doi:10.1002/nml.11203](https://doi.org/10.1002/nml.11203)
- Hoque, Z. (2004). A contingency model of the association between strategy, environmental uncertainty and performance measurement: Impact on organizational performance. *International Business Review*, *13*, 485-502. [doi:10.1016/j.ibusrev.2004.04.003](https://doi.org/10.1016/j.ibusrev.2004.04.003)
- Horsch, K. (1996). Results-based accountability systems: Opportunities and challenges. *The Evaluation Exchange*, *2*(1), 2-3.
- Hwang, H., & Powell, W. W. (2009). The rationalization of charity: The influences of professionalism in the nonprofit sector. *Administrative Science Quarterly*, *54*, 268-298. [doi: 10.2189/asqu.2009.54.2.268](https://doi.org/10.2189/asqu.2009.54.2.268)
- Jones, M. B. (2007). The multiple sources of mission drift. *Nonprofit and Voluntary Sector Quarterly*, *36*, 299-307. [doi:10.1177/0899764007300385](https://doi.org/10.1177/0899764007300385)
- Jordan, M. M., & Hackbart, M. M. (1999). Performance budgeting and performance funding in the states: A states assessment. *Public Budgeting and Finance*, *19*, 68-88. [doi:10.1046/j.0275-1100.1999.01157.x](https://doi.org/10.1046/j.0275-1100.1999.01157.x)
- Julian, D. A., & Kombarakaran, F. (2006). Assessment of quality of outcomes within a local united way organization: Implications for sustaining system level change. *American Journal of Community Psychology*, *38*, 175-181. [doi:10.1007/s10464-006-9072-1](https://doi.org/10.1007/s10464-006-9072-1)
- Julian, D., & Clapp, J. (2000). Outcomes based funding and the logic model: Mechanisms for coordinating human services at the local level. *Evaluation and Program Planning*, *23*(2), 231-240.
- Julnes, P. d. L., & Holzer, M. (2001). Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and

- implementation. *Public Administration Review*, 61, 693-708. [doi:10.1111/0033-3352.00140](https://doi.org/10.1111/0033-3352.00140)
- Karl, B. D. (1998). Volunteers and professionals: Many histories, many meanings. In W. W. Powell, & E. S. Clemens (Eds.), *Private action and the public good* (pp. 245-257). New Haven, CT: Yale University Press.
- Kim, S. (2005). Balancing competing accountability requirements: Challenges in performance improvement of the nonprofit human services agency. *Public Performance and Management Review*, 29(2), 145-163.
- Knickerbocker, F. T. (1973). Oligopolistic reaction and multinational enterprise. *The International Executive*, 15, 7-9. [doi:10.1002/tie.5060150205](https://doi.org/10.1002/tie.5060150205)
- Kumar, N., Scheer, L. K., & Steenkamp, J-B. E. M. (1995). The effects of supplier fairness on vulnerable resellers. *Journal of Marketing Research*, 32, 54-65. [doi:10.2307/3152110](https://doi.org/10.2307/3152110)
- Lee, C., & Nowell, B. (2015). A framework for assessing the performance of nonprofit organizations. *American Journal of Evaluation*, 36, 299-319. [doi:10.1177/1098214014545828](https://doi.org/10.1177/1098214014545828)
- Li, H., & Barry, J. (2012). *Interpreting interactions in logistic regression*. Ithaca, NY: Cornell Statistical Consulting Unit.
- Lieberman, M. B., & Asaba, S. (2006). Why do firms imitate each other? *Academy of Management Review*, 31, 366-385. [doi:10.5465/AMR.2006.20208686](https://doi.org/10.5465/AMR.2006.20208686)
- MacCrimmon, K., & Wehrung, D. (1986) *Taking risks: The management of uncertainty*. New York, NY: Free Press.
- MacIndoe, H., & Barman, E. (2012). How organizational stakeholders shape performance measurement in nonprofits: exploring a multidimensional measure. *Nonprofit and Voluntary Sector Quarterly*, 42, 716-738. [doi:10.1177/0899764012444351](https://doi.org/10.1177/0899764012444351)
- McLean, C., & Brouwer, C. (2010). *The effect of the economy on the nonprofit sector: A June 2010 survey*. Retrieved from <http://www.guidestar.org/ViewCmsFile.aspx?ContentID=2963>
- Melkers, J., & Willoughby, K. (2005). Models of performance measurement use in local governments: Understanding budgeting, communication, and lasting effects. *Public Administration Review*, 65, 180-190. [doi:10.1111/j.1540-6210.2005.00443.x](https://doi.org/10.1111/j.1540-6210.2005.00443.x)
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83, 340-363. [doi:10.1086/226550](https://doi.org/10.1086/226550)
- Mikkelsen, H. (2013). Universities and interpreter certification. *Translation and Interpreting*, 5, 66-78. [doi:10.12807/ti.105201.2013.a03](https://doi.org/10.12807/ti.105201.2013.a03)
- Mizruchi, M. S., & Fein, L. C. (1999). The social construction of organizational knowledge: A study of the uses of coercive, mimetic, and normative isomorphism. *Administrative Science Quarterly*, 44, 653-683. [doi:10.2307/2667051](https://doi.org/10.2307/2667051)
- Modell, S. (2009). Institutional research on performance measurement and management in the public sector accounting literature: a review and assessment. *Financial Accountability and Management*, 25, 277-303. [doi:10.1111/j.1468-0408.2009.00477.x](https://doi.org/10.1111/j.1468-0408.2009.00477.x)
- Moore, M. H. (2003). *The public value scorecard: a rejoinder and an alternative to "strategic performance measurement and management in non-profit organizations" by Robert Kaplan* (Working Paper No. 18). Cambridge, MA: Hauser Center for Nonprofit Organizations.
- Moxham, C. (2009). Performance measurement: Examining the applicability of the existing body of knowledge to nonprofit organisations. *International Journal of Operations and Production Management*, 29, 740-763. [doi:10.1108/01443570910971405](https://doi.org/10.1108/01443570910971405)
- Poole, D. L., Davis, J. K., Reisman, J., & Nelson, J. E. (2001). Improving the quality of outcome evaluation plans. *Nonprofit Management and Leadership*, 11, 405-421. [doi:10.1002/nml.11402](https://doi.org/10.1002/nml.11402)
- Proehl, R. A. (2001). *Organizational change in the human services*. Thousand Oaks, CA: Sage.

- Radaelli, C. M. (2000). Policy transfer in the European Union: Institutional isomorphism as a source of legitimacy. *Governance*, 13, 25-43. doi:10.1111/0952-1895.00122
- Roy, C., & Seguin, F. (2000). The institutionalization of efficiency-oriented approaches for public service improvement. *Public Productivity and Management Review*, 23, 449-468. doi:10.2307/3380563
- Saigal, A. (2008). *A study on the impact of information technology use on nonprofit organizations* (Unpublished master's thesis). Grand Valley State University, Grand Rapids, MI.
- Salamon, L. M., Sokolowski, S. W., & Sturza, D. (1999). *America's nonprofit sector: A primer* (Vol. 10). New York, NY: Foundation Center.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7, 147-177. doi:10.1037/1082-989X.7.2.147
- Shapiro, S. P. (2005). Agency theory. *Annual Review of Sociology*, 31, 263-284. doi:10.1146/annurev.soc.31.041304.122159
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20, 571-610. doi:10.5465/AMR.1995.9508080331
- Taylor-Powell, E., & Henert, E. (2008). *Developing a logic model: Teaching and training guide*. Madison, WI: University of Wisconsin-Extension Cooperative Extension.
- Thompson, F. J., & Riccucci, N. M. (1998). Reinventing government. *Annual Review of Political Science*, 1, 231-257. doi:10.1146/annurev.polisci.1.1.231
- Thomson, D. E. (2010). Exploring the role of funders' performance reporting mandates in nonprofit performance measurement. *Nonprofit and Voluntary Sector Quarterly*, 39, 611-629. doi:10.1177/0899764009360575
- Tolbert, P. S., & Zucker, L. G. (1983). Institutional sources of change in the formal structure of organizations: The diffusion of civil service reform, 1880-1935. *Administrative Science Quarterly*, 28, 22-39. doi:10.2307/2392383
- Tornow, W. W. (1993). Perceptions or reality: Is multi-perspective measurement a means or an end? *Human Resource Management*, 32, 221-229. doi:10.1002/hrm.3930320203
- Tuckman, H. P., & Chang, C. F. (1991). A methodology for measuring the financial vulnerability of charitable nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, 20, 445-460. doi:10.1177/089976409102000407
- Ullian, L., & Schink, A. B. (2012). Digital M: Professionalism and credentialing in the field of volunteer management. In T. D. Connors (Ed.), *The volunteer management handbook: Leadership strategies for success* (pp. M.1-M.19). Hoboken, NJ: John Wiley & Sons.
- Van Slyke, D. M. (2007). Agents or stewards: Using theory to understand the government-nonprofit social service contracting relationship. *Journal of Public Administration Research and Theory*, 17, 157-187. doi:10.1093/jopart/mul012
- Wayman, J. C. (2003). *Multiple imputation for missing data: What is it and how can I use it?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Weber, M. (1978). *Economy and society: An outline of interpretative sociology*. Berkeley, CA: University of California Press.
- Westphal, J. D., Gulati, R., & Shortell, S. M. (1997). Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. *Administrative Science Quarterly*, 42, 366-394. doi:10.2307/2393924
- Wiener, S. J., Kirsch, A. D., & McCormack, M. T. (2002). *Balancing the scales: Measuring the roles and contributions of nonprofit organizations and religious congregations*. Washington, DC: Independent Sector.
- W.K. Kellogg Foundation. (2004). *W.K. Kellogg Foundation logic model development guide*. Battle Creek, MI: W.K. Kellogg Foundation.

- Yang, K. (2009). Examining perceived honest performance reporting by public organizations: Bureaucratic politics and organizational practice. *Journal of Public Administration Research and Theory*, 19, 81-105. [doi:10.1093/jopart/mum042](https://doi.org/10.1093/jopart/mum042)
- Zimmerman, M. A., & Zeitz, G. J. (2002). Beyond survival: Achieving new venture growth by building legitimacy. *Academy of Management Review*, 27, 414-431. [doi:10.5465/AMR.2002.7389921](https://doi.org/10.5465/AMR.2002.7389921)
- Zimmermann, J. A. M., & Stevens, B. W. (2006). The use of performance measurement in South Carolina nonprofits. *Nonprofit Management and Leadership*, 16, 315-327. [doi:10.1002/nml.109](https://doi.org/10.1002/nml.109)
- Zucker, L. G. (1987). Institutional theories of organization. *Annual Review of Sociology*, 13, 443-464. [doi:10.1146/annurev.so.13.080187.002303](https://doi.org/10.1146/annurev.so.13.080187.002303)

Author Biographies

Chongmyoung Lee is an assistant professor at California State University, Los Angeles. His research interests include public and nonprofit management, inter-organizational collaboration, and performance measurement. His recent works have appeared in *American Journal of Evaluation* and *Journal of Health and Human Services Administration*.

Richard M. Clerkin is executive director of the Institute for Nonprofits and associate professor in the Department of Public Administration at North Carolina State University. His research interests focus broadly on the nonprofit sector. In particular, he studies motivations for public service and public benefiting activities. His is a coauthor of the leading public administration textbook, *Public Administration Understanding Management, Politics, and Law in the Public Sector*, and research has been published in journals such as *Public Administration Review*, *American Review of Public Administration*, *Armed Forces and Society*, *Nonprofit and Voluntary Sector Quarterly*, and *Nonprofit Management and Leadership*.